



Research article

A deficiency of prescriptive analytics—No perfect predicted value or predicted distribution exists

Shuaian Wang¹, Xuecheng Tian^{1,*}, Ran Yan¹, and Yannick Liu²

¹ Department of Logistics and Maritime Studies, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

² Faculty of Business, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

* **Correspondence:** Email: simontxcheng@163.com.

Abstract: Researchers and industrial practitioners are now interested in combining machine learning (ML) and operations research and management science to develop prescriptive analytics frameworks. By and large, a single value or a discrete distribution with a finite number of scenarios is predicted using an ML model with an unknown parameter; the value or distribution is then fed into an optimization model with the unknown parameter to prescribe an optimal decision. In this paper, we prove a deficiency of prescriptive analytics, i.e., that no perfect predicted value or perfect predicted distribution exists in some cases. To illustrate this phenomenon, we consider three different frameworks of prescriptive analytics, namely, the predict-then-optimize framework, smart predict-then-optimize framework and weighted sample average approximation (w-SAA) framework. For these three frameworks, we use examples to show that prescriptive analytics may not be able to prescribe a full-information optimal decision, i.e., the optimal decision under the assumption that the distribution of the unknown parameter is given. Based on this finding, for practical prescriptive analytics problems, we suggest comparing the prescribed results among different frameworks to determine the most appropriate one.

Keywords: prescriptive analytics; predict-then-optimize; smart predict-then-optimize; weighted sample average approximation

1. Introduction

With the advancement of ML technologies and the accessibility of rich data, recent studies in data-driven optimization have illustrated the advantages of using rich feature data to reduce uncertainty in decision-making problems with uncertain parameters and thus improve the decision making performance [1,2]. Following the formulation of Bertsimas and Kallus [1], consider an optimization problem with a given cost function $c(y; z)$, where $y \in Y \subset \mathbb{R}^{d_y}$ denotes the uncertain parameter vector affecting the value of the cost function and $z \in Z \subset \mathbb{R}^{d_z}$ denotes the decision vector. Assume that there is an observation of auxiliary data $x^0 \in X \subset \mathbb{R}^{d_x}$, which is associated with y and will be used to predict y . The optimization problem can be mathematically formulated as follows:

$$z^*(x^0) \in \arg \min_{z \in Z} E_y[c(y; z) | x = x^0]. \quad (1)$$

To solve Optimization problem (1), a historical dataset denoted as $\{(x^i, y^i)\}_{i=1}^n$ is made accessible, where $x^i \in X$ is the historical auxiliary data vector, and $y^i \in Y$ is the corresponding historical realization of y .

To solve Optimization problem (1), the traditional method is to first use $\{(x^i, y^i)\}_{i=1}^n$ to build an ML model and then use the ML model to predict the value of y from the new example x^0 [1, 3]. That is, for every x , the ML model generates a predicted y , as denoted by $\hat{y}(x)$. A commonly used loss function to train the ML model with a continuous prediction target is the mean squared error (MSE) loss, which is expressed as follows:

$$L_{MSE} = \frac{1}{n} \sum_{i=1}^n (y^i - \hat{y}(x^i))^2. \quad (2)$$

After obtaining the predicted $\hat{y}(x^0)$, we then solve

$$\min_{z \in Z} c(\hat{y}(x^0); z) \quad (3)$$

to prescribe an optimal decision. This method is generally termed a predict-then-optimize (PO) framework or a two-stage framework [3].

Nonetheless, a good prediction may not lead to a good decision [1]. This is because Eq (2) does not consider the impact of $\hat{y}(x)$ on the downstream optimization problem. As a result, a more natural and appropriate method is to plug the optimization problem into the training of the ML model. That is, instead of focusing on minimizing the prediction error, we train the ML model by minimizing the decision error. This method is generally termed a smart PO (SPO) framework, an end-to-end learning framework, or a decision-focused learning framework [3,4]. A commonly used loss function, for ML models under these frameworks, namely, SPO loss, can be formulated as follows:

$$L_{SPO} = \frac{1}{n} \sum_{i=1}^n [c(y^i; z(\hat{y}^i)) - c(y^i; z(y^i))], \quad (4)$$

where the first term in square brackets represents the cost derived from the decision $z(\hat{y}^i) \in \arg \min_{z \in Z} c(\hat{y}^i; z)$ and the second term in square brackets denotes the cost of the full-information optimal decision $z(y^i) \in \arg \min_{z \in Z} c(y^i; z)$. However, training ML models using SPO loss might be

impossible computationally because of the nonconvex and discontinuous characteristics of the SPO loss function for combinatorial optimization problems.

In a recent study, Bertsimas and Kallus [1] considered another approach to take the uncertainty of y into account when solving Eq (1). In their method, they claimed that, for many traditional ML methods, predicting y from a new example x^0 can take the following form:

$$\hat{y}(x^0) = \sum_{i=1}^n w(x^i, x^0) y^i, \quad (5)$$

where $w(x^i, x^0)$ measures the similarity (closeness) between Example x^i in the historical data and the new example x^0 ; and, its format depends on the ML model used (i.e., random forest, k -nearest neighbor (kNN)) [5]. For example, if we use a kNN model, $w(x^i, x^0) = 1/k$ if x^i is a kNN of x^0 , and $w(x^i, x^0) = 0$ otherwise. This is because $w(x^i, x^0)$ can be regarded as an approximation of the conditional distribution of y given $x = x^0$, i.e., the approximate distribution of y has n scenarios y^1, \dots, y^n with probabilities $w(x^1, x^0), \dots, w(x^n, x^0)$ (and it is very likely that some probabilities are 0, meaning that the approximate distribution of y has less than n scenarios). Bertsimas and Kallus [1] then used

$$\min_z \sum_{i=1}^n w(x^i, x^0) c(y^i; z) \quad (6)$$

as an approximation of Eq (1) [5]. This method has been formally termed as a weighted sample average approximation (w-SAA) framework by Notz and Pibernik [6] because it combines local predictive ML methods (i.e., kNN) and traditional techniques for data-driven optimization (i.e., SAA).

For these prescriptive analytics frameworks, the main aim is to predict a perfect value of y (i.e., using a PO framework and an SPO framework) or approximate a perfect conditional distribution of y (i.e., using a w-SAA framework) to minimize the uncertain cost $c(y; z)$ after observing $x = x^0$. In other words, we hope to prescribe a perfect decision that, ideally, ought to be the same as the full-information optimal decision $z^*(x^0)$ if we can predict a perfect $\hat{y}(x^0)$ or a perfect conditional distribution of y given x^0 . However, does the perfect predicted value or the perfect predicted distribution really exist? In this paper, we aim to answer this question by proving a surprising fact:

No perfect predicted value exists in PO and SPO frameworks, and no perfect predicted distribution exists in a w-SAA framework.

To illustrate this phenomenon, we first design two examples under the PO framework and the SPO framework in Section 2. We then design two examples under the w-SAA framework in Section 3 and show that the w-SAA framework may perform better than the PO/SPO framework. At last, Section 4 concludes this paper.

2. Deficiency of PO and SPO frameworks

For most uncertain optimization problems with auxiliary data, a commonly used PO framework entails first predicting the value of uncertain parameters and then plugging predictions into the optimization problem to derive optimal decisions. As this method does not consider the impact of predictions on the downstream decisions, it may lead to sub-optimal decisions. Hence, Elmachtoub and Grigas [3] proposed an SPO framework for a broad class of decision-making problems with

uncertainty to deal with this issue. Either under the PO framework or the SPO framework, the aim is to predict a perfect uncertain parameter with the use of auxiliary data, so as to prescribe a full-information optimal decision if the perfect full information (i.e., the marginal distribution) of the uncertain parameter is known. However, it is possible that no perfect predicted value exists in PO and SPO frameworks, which shall be illustrated using two examples. We first show that this deficiency exists in a classification task under the PO/SPO framework.

Example 1. Consider that we have an odd number n of historical records (examples) in the auxiliary dataset $\{(x^i, y^i)\}_{i=1}^n$. Suppose that the features of all examples are identical (i.e., $x^1 = \dots = x^n$). For example, x^i indicates whether it is a work day, where $x^i = 1$ indicates that it is a work day and $x^i = 0$ otherwise, thus, $x^i \in \{0, 1\}$. Additionally, suppose that the uncertain parameter $y \in \{0, 1\}$ follows a discrete uniform distribution independent of x and that the optimization problem with the cost function $c(y; z)$ is as follows:

$$\min_{-1 \leq z_1, z_2 \leq 1} [(y - 0.7)z_1 + (y - 0.3)z_2], \quad (7)$$

where $z = (z_1, z_2)$ is the decision vector.

Suppose $x^0 = x^1 = \dots = x^n$. After observing x^0 , the resulting decision problem is

$$\min_{-1 \leq z_1, z_2 \leq 1} E_y[(y - 0.7)z_1 + (y - 0.3)z_2 \mid x = x^0]. \quad (8)$$

If we know in advance that the uncertain parameter y follows a discrete uniform distribution with two values 0 and 1 independent of x , the above model can be written as:

$$\min_{-1 \leq z_1, z_2 \leq 1} \{0.5[(1 - 0.7)z_1 + (1 - 0.3)z_2] + 0.5[(0 - 0.7)z_1 + (0 - 0.3)z_2]\}. \quad (9)$$

Therefore, the unique full-information optimal decision for Optimization problem (8) is $z^*(x^0) = (z_1^*(x^0), z_2^*(x^0)) = (1, -1)$.

Suppose we use a classification ML model to predict the value of y (0 or 1) given the new example x^0 . Since we have n records satisfying $x^0 = x^1 = \dots = x^n$ and n is an odd number, if $\sum_{i=1}^n I(y^i = 1) > \sum_{i=1}^n I(y^i = 0)$ (where $I(q)$ is an indicator function which takes the value of 1 if Condition q is true and 0 if Condition q is false), we can obtain $\hat{y} = 1$ for the new example x^0 by using a classification ML model. This is because predicting y to be 1 for all examples in the historical dataset can minimize Loss function (2) under the PO framework and minimize Loss function (4) under the SPO framework. Given that $\hat{y} = 1$, Optimization model (7) can be written as:

$$\min_{-1 \leq z_1, z_2 \leq 1} [(1 - 0.7)z_1 + (1 - 0.3)z_2], \quad (10)$$

and the prescribed optimal decision is $z^*(\hat{y}) = (z_1^*(\hat{y}), z_2^*(\hat{y})) = (-1, -1)$. Similarly, if $\sum_{i=1}^n I(y_i = 0) > \sum_{i=1}^n I(y_i = 1)$, we can obtain $\hat{y} = 0$ for the new example x^0 . Given that $\hat{y} = 0$, Optimization model (7) can be written as

$$\min_{-1 \leq z_1, z_2 \leq 1} [(0 - 0.7)z_1 + (0 - 0.3)z_2], \quad (11)$$

and the prescribed optimal decision is $z^*(\hat{y}) = (z_1^*(\hat{y}), z_2^*(\hat{y})) = (1, 1)$. So, for this optimization problem, $z^*(x^0) \neq z^*(\hat{y})$, which means that we can never predict a perfect \hat{y} to prescribe the full-information optimal decision $z^*(x^0) = (1, -1)$.

From Example 1, we find that under the PO/SPO framework, we may not be able to predict a perfect value to prescribe a full-information optimal decision with a classification ML model. This phenomenon can also be seen in a regression task, as illustrated in the following example.

Example 2. Suppose that we have an odd number n of historical records (examples) in the auxiliary dataset $\{(x^i, y^i)\}_{i=1}^n$. Suppose that the features of all examples are identical (i.e., $x^1 = \dots = x^n$). For example, x^i indicates whether it is a work day and $x^i \in \{0, 1\}$. Additionally, suppose that the uncertain parameter y is a real number following a uniform distribution independent of x with the following probability density function:

$$f(y) = \frac{1}{2b}, -b < y < b, \quad (12)$$

where b is a positive real value. The optimization problem with the cost function $c(y; z)$ is established as follows:

$$\min_{-1 \leq z_1, z_2 \leq 1} [(I(y > 0) - 0.7)z_1 + (I(y > 0) - 0.3)z_2], \quad (13)$$

where $z = (z_1, z_2)$ is the decision vector.

Suppose that $x^0 = x^1 = \dots = x^n$. After observing x^0 , the resulting decision problem is

$$\min_{-1 \leq z_1, z_2 \leq 1} E_y [(I(y > 0) - 0.7)z_1 + (I(y > 0) - 0.3)z_2 | x = x^0]. \quad (14)$$

If we know in advance that the uncertain parameter y follows a uniform distribution independent of x and that its probability density function is as shown in Eq (12), the above model can be written as:

$$\min_{-1 \leq z_1, z_2 \leq 1} 0.5[(1 - 0.7)z_1 + (1 - 0.3)z_2] + 0.5[(0 - 0.7)z_1 + (0 - 0.3)z_2], \quad (15)$$

and the unique full-information optimal decision for Optimization problem (14) is $z^*(x^0) = (z_1^*(x^0), z_2^*(x^0)) = (1, -1)$.

Suppose we use a regression ML model to predict the value of y (a real number) given the new example x^0 . Since we have n records satisfying $x^0 = x^1 = \dots = x^n$ and n is an odd number, if $\sum_{i=1}^n I(y_i > 0) > \sum_{i=1}^n I(y_i \leq 0)$, we can obtain $\hat{y} > 0$ for the new example x^0 by using a regression ML model. This is because predicting y to be bigger than 0 for all examples in the historical dataset can minimize Loss function (2) under the PO framework and minimize Loss function (4) under the SPO framework. Given that $\hat{y} > 0$, Optimization model (7) can be written as:

$$\min_{-1 \leq z_1, z_2 \leq 1} [(1 - 0.7)z_1 + (1 - 0.3)z_2], \quad (16)$$

and the prescribed optimal decision is $z^*(\hat{y}) = (z_1^*(\hat{y}), z_2^*(\hat{y})) = (-1, -1)$. Similarly, if

$\sum_{i=1}^n I(y_i \leq 0) > \sum_{i=1}^n I(y_i > 0)$, we can obtain $\hat{y} \leq 0$ for the new example x^0 . Given that $\hat{y} = 0$,

Optimization model (7) can be written as:

$$\min_{-1 \leq z_1, z_2 \leq 1} [(0 - 0.7)z_1 + (0 - 0.3)z_2], \quad (17)$$

and the prescribed optimal decision is $z^*(\hat{y}) = (z_1^*(\hat{y}), z_2^*(\hat{y})) = (1, 1)$. So, for this optimization problem, $z^*(x^0) \neq z^*(\hat{y})$, which means that we can never predict a perfect \hat{y} to prescribe the full-information optimal decision $z^*(x^0) = (1, -1)$.

From Examples 1 and 2, we have proved that, under the PO/SPO framework for both regression and classification tasks, there may not be a perfect predicted value that will prescribe the full-information optimal solution. There are, of course, situations in which there is a perfect predicted value that will prescribe a full-information optimal solution, which is shown in Proposition 1, as follows.

Proposition 1. If $c(y; z)$ is linear in y , then there exists a perfect predicted value that will prescribe a full-information optimal solution if y can take any value in \mathbb{R}^d in the predictive ML model.

Proof: If $c(y; z)$ is linear in y , the model given by Eq (1) can be represented by $\min_{z \in Z} E_y[c(y; z) | x = x^0] = \min_{z \in Z} E_y[y | x = x^0]z$. Hence, $\hat{y} = E[y | x = x^0]$ can lead to the full-information optimal decision.

If there exists a perfect predicted value that will prescribe a full-information optimal solution, the perfect predicted value is not necessarily the conditional mean value, which is shown in Proposition 2, as follows.

Proposition 2. In the traditional newsvendor problem with a continuous order quantity (denoted by z) and demand quantity (denoted by y) [7,8],

$$c(y; z) = o \max(z - y, 0) + u \max(y - z, 0) \quad (18)$$

is the random cost, where o and u are the overage cost and underage cost, respectively. If the marginal demand distribution of y (with a given x^0 , e.g., whether it is a workday) is known in advance and is denoted as F_{x^0} , the full-information optimal decision is $z^*(x^0) = F_{x^0}^{-1}(u / (u + o))$. Hence, the predicted value $\hat{y} = F_{x^0}^{-1}(u / (u + o))$ will prescribe a full-information optimal solution. Furthermore, in the traditional newsvendor problem with an integer demand quantity and integer order quantity, the predicted value \hat{y} , which is the smallest integer satisfying $F_{x^0}(\hat{y}) \geq u / (u + o)$, will prescribe a full-information optimal solution.

3. Deficiency of w-SAA method

In Section 2, we presented two examples to show the deficiency of PO and SPO frameworks, demonstrating that it might be impossible to predict a perfect value to prescribe a full-information optimal decision. A similar phenomenon also exists in the w-SAA framework. The w-SAA framework was initially proposed by Bertsimas and Kallus [1], and it is based on the idea of deriving weights from features by developing local predictive ML methods and optimizing the decision against a reweighting of the data [6]. According to the format of Eq (5), this method aims to estimate the conditional distribution of y given the new example x^0 . Taking the kNN method as an example, we assume that the predicted target \hat{y} follows a discrete uniform distribution with no more than K values (neighbors). Here, K is not an infinite positive number. However, with full information, the real y may either follow a discrete distribution with more than K values, or a continuous distribution. Then,

given the above information, we use the following example to show that a deficiency of the w-SAA framework also exists, that is, no perfect predicted distribution exists under the w-SAA framework.

Example 3. Consider that we have an auxiliary dataset $\{(x^i, y^i)\}_{i=1}^n$. Suppose that the features of all historical data are identical (i.e., $x^1 = \dots = x^n$). For example, x^i indicates whether it is a work day and $x^i \in \{0, 1\}$. Additionally, suppose that the uncertain parameter y is a real number following a discrete uniform distribution independent of x and $y \in \{1, 2, 3, \dots, K+1\}$. Let $0 < \varepsilon < 1$, which is a parameter used to formulate the following optimization problem. The optimization problem with the cost function $c(y; z)$ is established as follows:

$$\min_{-1 \leq z_{k,1}, z_{k,2} \leq 1} \sum_{k=1, \dots, K+1} \left[\left(\mathbb{I}(y=k) - \frac{(1+\varepsilon)}{K+1} \right) z_{k,1} + \left(\mathbb{I}(y=k) - \frac{(1-\varepsilon)}{K+1} \right) z_{k,2} \right]. \quad (19)$$

Suppose that $x^0 = x^1 = \dots = x^n$. After observing x^0 , the resulting decision problem is

$$\min_{-1 \leq z_{k,1}, z_{k,2} \leq 1} \mathbb{E}_y \left\{ \sum_{k=1, \dots, K+1} \left[\left(\mathbb{I}(y=k) - \frac{(1+\varepsilon)}{K+1} \right) z_{k,1} + \left(\mathbb{I}(y=k) - \frac{(1-\varepsilon)}{K+1} \right) z_{k,2} \mid x = x^0 \right] \right\}. \quad (20)$$

If we know in advance that the uncertain parameter y follows a discrete uniform distribution, the above model can be written as:

$$\min_{-1 \leq z_{k,1}, z_{k,2} \leq 1} \sum_{k=1, \dots, K+1} \left\{ \begin{array}{l} \frac{1}{K+1} \left(1 - \frac{(1+\varepsilon)}{K+1} \right) z_{k,1} + \frac{K}{K+1} \left(0 - \frac{(1+\varepsilon)}{K+1} \right) z_{k,1} \\ + \frac{1}{K+1} \left(1 - \frac{(1-\varepsilon)}{K+1} \right) z_{k,2} + \frac{K}{K+1} \left(0 - \frac{(1-\varepsilon)}{K+1} \right) z_{k,2} \end{array} \right\}, \quad (21)$$

and the unique full-information optimal decision for Optimization problem (20) is $z_{k,1}^*(x^0) = 1, \forall k = 1, \dots, K+1$ and $z_{k,2}^*(x^0) = -1, \forall k = 1, \dots, K+1$.

Suppose we use the w-SAA framework with a kNN model to predict a conditional distribution of y with no more than K values given the new example x^0 . These values are denoted by $y_1, \dots, y_{k'}, \dots, y_K$ and each value has a weight $w(y_{k'}) = 1/K, \forall k' = 1, \dots, K$. For other data examples that do not have a target value in $\{y_1, \dots, y_{k'}, \dots, y_K\}$, they will be assigned a weight of 0. In order to obtain the full-information optimal decision, the following condition must be satisfied when we solve Optimization problem (6): $\forall k = 1, \dots, K+1, \frac{1-\varepsilon}{K+1} \leq \sum_{k' \in \{y_{k'}=k\}} w(y_{k'}) \leq \frac{1+\varepsilon}{K+1}$. However, since $y_{k'}$ only

has no more than K values, the above condition cannot be met because $\exists k = 1, \dots, K+1, \sum_{k' \in \{y_{k'}=k\}} w(y_{k'}) = 0$. That is, we are not able to prescribe a decision, which, ideally,

ought to be the same as the full-information optimal decision in the w-SAA framework. Similarly, this phenomenon can be extended when the uncertain parameter y follows a continuous distribution.

Note that the prediction of a single value in the PO/SPO framework can be considered as a special case of the w-SAA framework that involves applying a single scenario to approximate the marginal distribution. Based on this insight, the following Example 4 shows that the w-SAA framework can prescribe better solutions than the PO/SPO framework.

Example 4. (Example 3 continued) Suppose we set $K = 2$ for Example 3, that is, the uncertain parameter y is a real number following a discrete uniform distribution where $y \in \{1, 2, 3\}$; then, we

conduct a three-class classification task. Suppose that $x^0 = x^1 = \dots = x^n$. After observing x^0 , the full-information optimal decision for Optimization problem (19) is $z_{k,1}^*(x^0) = 1, \forall k = 1, 2, 3$ and $z_{k,2}^*(x^0) = -1, \forall k = 1, 2, 3$.

Let us use the w-SAA framework with a kNN model to find the nearest two examples of the new example x^0 denoted by $\{(x_j, y_j)\}_{j=1}^2$. These two examples are assigned a weight of $w_i = 1/2, \forall i = 1, 2$, and the rest of the examples $(x_j, y_j), \forall j = 3, \dots, n$ are assigned a weight of 0. There are six possible combinations for (y_1, y_2) , which are (1,1), (2,2), (3,3), (1,2), (1,3) and (2,3). For the first three combinations, following the same procedures shown in Example 3, the prescribed cost will be 0. Similarly, for the last three combinations, the prescribed cost will be $-(4\varepsilon/3)$.

If we use a PO/SPO framework to predict the value of y , there are three possible outcomes, i.e., $\hat{y} = 1, \hat{y} = 2$, and $\hat{y} = 3$. No matter under which outcome, if we plug the prediction into Optimization problem (19), the prescribed cost will be 0. Since $0 \leq 0$ and $-(4\varepsilon/3) < 0$, it shows that the w-SAA framework can prescribe a better solution.

Note further that the w-SAA framework approximates the conditional distribution of y in a local manner (i.e., by using a portion of the training data). Wang and Yan [9,10] proposed two methods to approximate the conditional distribution of y in a global manner (i.e., by using all of the training data). Similarly, we can show that there are cases for which no perfect predicted distribution exists under the conditions of the methods proposed by Wang and Yan [9,10].

From the above examples, we have shown that it is possible that prescriptive analytics cannot predict a perfect value or a perfect discrete distribution for finite scenarios. Nonetheless, different frameworks may prescribe solutions with different qualities, and our aim is to find a better framework that can prescribe a near-optimal solution.

4. Conclusions

Using abundant auxiliary data and ML models, researchers and industrial practitioners can make better decisions. However, using current prescriptive analytics frameworks, we may not be able to prescribe a full-information optimal decision. This paper thus makes the following contributions. First, this paper proves a deficiency of prescriptive analytics, i.e., that no perfect predicted value or predicted distribution exists; this was demonstrated by presenting three examples under different frameworks, namely the PO framework, SPO framework, and w-SAA framework. From these three examples, we show that prescriptive analytics may not be able to prescribe a perfect decision which ideally ought to be the same as the full-information optimal decision.

Second, this paper inspires researchers and practitioners to check the existence of the perfect predicted value or the perfect predicted distribution while using prescriptive analytics frameworks. If there is no perfect predicted value or predicted distribution for an uncertain optimization problem, we may need to try more than one framework and compare the prescribed results obtained from different frameworks. As indicated by Example 4, it is possible for us to prescribe a better solution by using a w-SAA framework rather than a PO/SPO framework. To summarize, our research demonstrates that neither PO/SPO nor w-SAA is perfect, and that, consequently, more efforts should be devoted to the field of prescriptive analytics.

There are several limitations to this research. First, this paper only includes a theoretical analysis, as the available practical data are insufficient. In the future, practical problems with real data can be

used to verify the deficiency demonstrated in this paper. Second, when analyzing the superiority of different prescriptive analytics frameworks, it should be taken into consideration that this paper does not prove a theoretical gap between their decision-making qualities, which should be explored further in the future.

Acknowledgments

The authors thank the two reviewers for their valuable comments.

Conflict of interest

The authors declare that there is no conflict of interest.

References

1. D. Bertsimas, N. Kallus, From predictive to prescriptive analytics. *Manage. Sci.*, **66** (2020), 1025–1044. <https://doi.org/10.1287/mnsc.2018.3253>
2. T. Olovsson, T. Svensson, J. Wu, Future connected vehicles: Communications demands, privacy and cyber-security, *Commun. Transp. Res.*, **2** (2022), 100056. <https://doi.org/10.1016/j.commtr.2022.100056>
3. A. N. Elmachtoub, P. Grigas, Smart “predict, then optimize”, *Manage. Sci.*, **68** (2021), 9–26. <https://doi.org/10.1287/mnsc.2020.3922>
4. M. Mulamba, J. Mandi, M. Diligenti, M. Lombardi, V. Bucarey, T. Guns, Contrastive losses and solution caching for predict-then-optimize, in *Proceedings of 2021 International Joint Conference on Artificial Intelligence*, (2021), 2833–2840. <https://arxiv.org/abs/2011.05354v2>
5. D. Bertsimas, N. Koduri, Data-driven optimization: A reproducing kernel Hilbert space approach, *Oper. Res.*, **70** (2021), 454–471. <https://doi.org/10.1287/opre.2020.2069>
6. P. Notz, R. Pibernik, Prescriptive analytics for flexible capacity management, *Manage. Sci.*, **68** (2022), 1756–1775. <https://doi.org/10.1287/mnsc.2020.3867>
7. L. Chen, D. Long, G. Perakis, The impact of a target on newsvendor decisions, *Manuf. Serv. Oper. Manage.*, **17** (2015), 78–86. <https://doi.org/10.1287/msom.2014.0500>
8. G. Y. Ban, C. Rudin, The big data newsvendor: Practical insights from machine learning, *Oper. Res.*, **67** (2019), 90–108. <https://doi.org/10.1287/opre.2018.1757>
9. S. Wang, R. Yan, A global method from predictive to prescriptive analytics considering prediction error for “Predict, then optimize” with an example of low-carbon logistics, *Cleaner Log. Supply Chain*, **4** (2022), 100062. <https://doi.org/10.1016/j.clscn.2022.100062>
10. S. Wang, R. Yan, “Predict, then optimize” with quantile regression: A global method from predictive to prescriptive analytics and applications to transportation, *Multi. Transp.*, (2022), in press.



AIMS Press

©2022 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)