Multimodal Transportation 1 (2022) 100018

Contents lists available at ScienceDirect

Multimodal Transportation

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

Editorial

Integrating prediction with optimization: Models and applications in transportation management

Ran Yan, Shuaian Wang*

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

ARTICLE INFO

Keywords:
Prediction
optimization
predict-then-optimize
smart “predict-then-optimize”
predictive prescription

ABSTRACT

Prediction and optimization are the foundation of many real-world analytics problems in various disciplines. As both can be challenging, they are usually treated sequentially in existing studies, where the prediction problem is dealt with in the first stage, followed by the optimization problem in the second stage, which is called the predict-then-optimize paradigm. Specifically, the unknown parameters in the optimization problem are first predicted by the prediction model and are then input to the optimization model to generate the optimal decisions. However, prediction models in the first stage are intended to minimize the prediction error, while ignoring the structure and property of the downstream optimization problem and how the predictions will be used. Consequently, suboptimal decisions might be generated. This editorial piece discusses current popular frameworks to integrate prediction with optimization, namely the smart “predict, then optimize” framework and the predictive prescription framework with examples in the transportation area provided. The article ends with proposing several promising research directions for future research.

1. Introduction

Prediction and optimization are two forefronts for decision making and planning in modern business analytics such as transportation management (Peng et al., 2021; Wang and Wu, 2021). Prediction is based on the premise that current and past knowledge is useful to predict the future by extracting patterns from historical data and implementing them to predict future values. Among various prediction approaches, statistical modelling is a classic one based on statistics theory. With rapid advances in computing, the prediction’s toolbox of methods has grown in size and sophistication, enabling the analysis of larger and more complex datasets. ML, as a subfield of artificial intelligence (AI), is a representative popular emerging approach for prediction tasks due to its high prediction accuracy. A recent review on the application of AI technologies to address practical problems in transportation is conducted by Abduljabbar et al. (2019). Prediction models discussed in the current study mainly refer to those based on ML models. An optimization problem aims to find the optimal solution to a mathematical model which is an abstraction of a practical problem by minimizing or maximizing the value of an objective function (or a set of objective functions) while satisfying a set of constraints. There are several classes of optimization models considering variable values (continuous optimization or discrete optimization), objective functions (none, one, or many objectives), model constraints (unconstrained optimization or constrained optimization), and randomness (deterministic optimization or stochastic optimization).

Most solution systems for real-world analytics problems involve some component of both prediction and optimization approaches, with data as the base (Elmachtoub and Grigas, 2021). Considering the complex nature of both approaches, a standard paradigm is

* Corresponding author at: The Hong Kong Polytechnic University, Department of Logistics and Maritime Studies, Hung Hom, Kowloon, Hong Kong, China

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

https://doi.org/10.1016/j.multra.2022.100018
Received 17 November 2021; Received in revised form 26 December 2021; Accepted 5 January 2022

to develop a sequential two-stage framework named predict-then-optimize, which is central to many applications of ML for use in operations research problems. Particularly, key unknown parameters of the optimization model are first predicted by an ML model, whose aim is to minimize the prediction error evaluated by metrics of prediction tasks such as mean squared error (MSE) and mean absolute error (MAE). Then, the downstream optimization model where the unknown parameters are predicted by the ML model is solved. One major challenge in this process is that the parameters cannot be perfectly predicted due to the factors not considered and the limited precision of measurements in any ML model. Consequently, such errors are inevitably brought to the following optimization model and thus causing inefficiency in decision making. Furthermore, as traditional ML models totally ignore how the predictions will be used in and interact with the following optimization model while focusing solely on minimizing prediction errors in the construction process and evaluation metrics, the effectiveness of the predict-then-optimize framework might be further weakened. We use the following simple and intuitive example in the transportation area to illustrate this issue.

1.1. Example I. Shortest-path problem

Suppose Mary plans to drive from an origin A to a destination B using the least amount of time, and there are two alternative paths, namely path 1 and path 2, that can be chosen from. Although the travel time on both paths is unknown before traveling, it can be predicted by auxiliary data such as total distance, congestion condition, speed limit, weather, and time of day. Suppose further that the real travel time on path 1 and path 2 is 10 min and 20 min, respectively, and thus the optimal choice under perfect information should be path 1. Now we have two ML models, and the travel time on path 1 is predicted to be 16 min and 20 min, and that on path 2 is predicted to be 14 min and 30 min by ML model 1 and ML model 2, respectively. Therefore, path 2 should be selected according to ML model 1, and path 1 should be selected according to ML model 2. It can easily be seen that although the prediction accuracy of ML model 1 is much higher than that of ML model 2, it generates a worse decision.

The above example clearly shows that in the traditional two-stage predict-then-optimize framework, a better ML model does not necessarily lead to a better decision (Bertsimas and Kallus, 2020; Elmachtoub and Grigas, 2021). This finding might be contrary to our intuition: fundamentally, prediction models with higher accuracy would induce better decisions. The overall working flow of this framework indicates that even if the predict-then-optimize approach consists of a “prediction” stage and an “optimization” stage in a unified framework, they seem to be separated parts as they are implemented sequentially with rare interactions. Therefore, a natural question is that as better prediction does not necessarily lead to better decisions, how to integrate prediction with optimization more effectively to generate more informed decisions? In the existing literature, two types of approaches are receiving wide attention: smart “predict, then optimize” (SPO) framework, which directly considers the decision error when developing prediction models, and predictive prescription, which considers the joint distribution of auxiliary data and the unknown parameters in the downstream optimization model. Detailed methodology and representative literature are provided in the next section.

2. Two emerging frameworks integrating prediction with optimization

2.1. SPO framework

The SPO framework is formally proposed by Elmachtoub and Grigas (2021), which is designed to be implemented in optimization models with a linear objective with unknown parameters subject to a convex feasible region. It directly leverages the structure of the optimization model (including its objective and constraints) in the second stage to design a better prediction model in the first stage, which is quite intuitive. In brief, the SPO framework directly integrates prediction with optimization, where the decision error induced by predictions generated is used to train the prediction model instead of the prediction error itself. Therefore, the core of the SPO framework is the SPO loss function. Mathematically, in *Example 1*, define $\mathbf{c}$ as the vector of real travel time on each alternative path, and $\mathbf{z}$ as a binary decision vector representing whether each alternative path is traversed. Given the vector of the predicted travel time on each alternative path $\hat{\mathbf{c}}$, the optimal solution can be calculated by $\mathbf{z}^*(\hat{\mathbf{c}}) = \min_{z \in S} \hat{\mathbf{c}}^T \mathbf{z}$, where S is the feasible region. The actually realized solution (based on real travel time) given $\mathbf{z}^*(\hat{\mathbf{c}})$ is $\mathbf{c}^T \mathbf{z}^*(\hat{\mathbf{c}})$. The main difference between the traditional predict-then-optimize framework and the SPO framework can be illustrated by the different evaluation metrics of prediction models: given two ML models $f^{(1)}$ and $f^{(2)}$ with $\mathbf{x}$ as the set of features as the input, if $\frac{1}{N}\sum_{i=1}^{N} \left(f^{(1)}(\mathbf{x}^i) - \mathbf{c}^i\right)^2 < \frac{1}{N}\sum_{i=1}^{N} \left(f^{(2)}(\mathbf{x}^i) - \mathbf{c}^i\right)^2$, i.e. $f^{(1)}$ can generate more accurate predictions than $f^{(2)}$, $f^{(1)}$, is regarded to be better than $f^{(2)}$ under the standard predict-then-optimize framework. In contrast, if $\frac{1}{N}\sum_{i=1}^{N} \mathbf{c}^{iT}\mathbf{z}^*(f^{(1)}(\mathbf{x}^i)) < \frac{1}{N}\sum_{i=1}^{N} \mathbf{c}^{iT}\mathbf{z}^*(f^{(2)}(\mathbf{x}^i))$, i.e. $f^{(1)}$ can induce better decision than $f^{(2)}$, $f^{(1)}$, is regarded to be better than $f^{(2)}$ under the SPO framework. It should also be mentioned that which of the two approaches performs better depends highly on the specific problem and should be evaluated by a separate test set (which is the so called “no free lunch theorem”).

It is also pointed out by Elmachtoub and Grigas (2021) that training a prediction model with respect to the SPO loss is computationally challenging, as it may require solving the downstream optimization model thousands of times. Therefore, a tractable convex surrogate loss function called SPO+ loss is derived from the original SPO loss based on duality theory. The authors also prove that the SPO+ loss is statistically consistent with the SPO loss under mild conditions, while it can relieve the computational burden to a large extent (Elmachtoub and Grigas, 2021). Similarly, differentiable surrogate loss functions are employed in Wilder et al. (2019), Mandi et al. (2020), and Ferber et al. (2020) to accelerate the solution of the optimization model based on the idea of the SPO framework.

We would like to add more comments on the superior performance of the SPO framework over standard paradigm. The major reason is that the SPO framework integrates prediction with optimization. In the traditional predict-then-optimize framework, the prediction model does not account for how the prediction will be used in the following optimization model. The target of the prediction model is to minimize the prediction error, that is, the difference between the predicted and the real output values. Consequently, as is shown in *Example 1*, better prediction would not necessarily lead to better decision. In the SPO framework, the prediction model directly leverages the structure and the constraints of the optimization model and adopts a tailored loss function accordingly. The target is thus to minimize the decision error (cost) generated in the following optimization model. This is in essence analogous to the standard predict-then-optimize framework, while the main difference is that the former aims to minimize the decision error, and the latter aims to minimize the prediction error, and such difference is reflected by the adoption of different loss functions. One can understand the prediction given by the SPO framework as it aims to generate 'personalized' prediction (which is not necessarily more accurate) suitable for improving the solution generated by the downstream optimization model. As the standard predict-then-optimize framework can provide better prediction than random guess, it can also be expected that the SPO framework is more likely to generate better decisions than the standard predict-then-optimize framework.

Since the SPO framework was proposed, there are many studies aiming to expand its theory and application. The worst-case generalization bounds of the SPO loss are derived by using combinatory parameters measuring the capacity of function classes, and its margin-based generalization bounds are derived through exploiting the special structure in data (Balghiti et al., 2019). A tractable methodology called SPO Trees (SPOTs) is proposed by Elmachtoub (2020) to train decision tree models under the SPO loss. Moreover, to facilitate the construction of random forest (RF) models with SPO loss, approximate splitting criteria based on optimization perturbation analysis is developed to relieve the computational burden when finding the optimal split for each node, and thus allows solving large-scale problems (Kallus and Mao, 2021). Furthermore, the SPO loss is directly minimized in Demirović (2019) by learning linear functions for ranking objectives and in Demirović (2020) by applying dynamic programming to solve optimization problem with linear learning functions. The SPO loss function is further extended by decision-driven regularization in Loke et al. (2021). In addition to theoretical analysis, the SPO framework is applied to solve practical problems, such as last-mile delivery planning for an online food delivery platform (Chu et al., 2021) and ship selection for inspection in port state control (PSC) (Yan et al., 2020).

We also observe that the semi-SPO framework, which only considers the property and structure of the downstream optimization model instead of solving the optimization model itself and evaluating the decision quality it generates when training the first stage ML model, is proposed by existing studies. The semi-SPO framework partly refers to the basic idea of the SPO framework for integrating prediction and optimization while lying in-between the standard predict-then-optimize paradigm and the SPO framework. A typical example is Yan et al. (2020), where in the first stage a multi-target regression model is developed to predict the risk level under each risk category of a ship, and the second stage is a PSC inspector assignment model to match the inspectors' expertise with the predicted ship risk condition, so as to maximize the total amount of ship risk that can be identified given certain inspection resources. Mathematically, given a set of ships to be inspected denoted by S and a set of PSC inspectors by P, the risk that can be identified if assigning inspector $p \in P$ to inspect ship $s \in S$ is denoted by v_{ps}. In the traditional predict-then-optimize paradigm, the loss function taking the form of $\frac{1}{|S|}\sum_{s\in S}\sum_{p\in P}(\hat{v}_{ps}-v_{ps})^2$ is to be minimized when training the multi-target prediction model in the first stage, which aims to minimize the MSE. After examining the structure and property of the assignment model, it can easily be found that for a ship $s \in S$, if all v_{ps}, $p \in P$ are overestimated or underestimated by the same amount, there is no effect on the optimal assignment. This is because for a certain ship, it is the relative relationship of v_{ps} where $p \in P$, instead of their absolute values, that will influence the optimal decision. Therefore, motived by this observation, a semi-SPO approach is developed by the authors that minimizes loss function $\frac{1}{|S|}\sum_{s\in S}\sum_{p\in P}\sum_{p'\in P\setminus\{p\}}[(\hat{v}_{ps}-v_{ps})-(\hat{v}_{p's}-v_{p's})]^2$ in the first stage ML model, which aims to minimize the sum of the differences in overestimates in the numbers of deficiencies that can be detected among the PSCOs for each pair of ships.

2.2. Predictive prescription framework

Unlike the SPO framework which is applied to optimization models with linear objective functions in the unknown parameters, the predictive prescription framework is applied when the objective function is nonlinear. Consider a case of the predict-then-optimize paradigm where there are unknown parameters in the downstream objective function. Suppose that we have infinite historical data on the unknown parameter, and thus its mean value can be accurately estimated, which is the best estimate of an ML model that minimizes the MSE. If the objective function is linear in the unknown parameter, the optimal solutions can be found in this case thanks to the linearity. Nevertheless, if the objective function is nonlinear in the unknown parameters, feeding their mean values into the optimization model is likely to produce sub-optimal solutions due to the uncertainty brought by the nonlinearity in model parameters. Furthermore, in reality, we only have limited data to estimate the distributions as well as the mean values of the unknown parameters (Dragomir and Dumitru, 2022; Schinas and Sonechko, 2022), and hence the drawback of the predict-then-optimize paradigm may be even more significant due to inaccurate estimation.

To address the issues brought by the nonlinearity of the unknown parameters in the objective function, a better paradigm is proposed by Bertsimas and Kallus (2020) to estimate their distributions by leveraging auxiliary observations, which the authors refer to as predictive prescription. We would like to first clarify why predictive prescription is only to be applied when we have a nonlinear problem where the objective function is nonlinear in the unknown parameters with the reason as follows. Under the condition that we have infinitely many data, the mean value of the unknown parameter is thus the best estimate in an ML model that minimizes the MSE. Given the condition that the objective function is linear in the unknown parameters, the expectation operator on the unknown parameters can be moved to operate the whole objective function, and thus the expectation of the whole objective function can be minimized or maximized. On the contrary, if the objective function is nonlinear in the unknown parameters, even if the expectation of the unknown parameters can be obtained, such expectation cannot be applied to the expectation of the whole objective function.

To address this issue, Bertsimas and Kallus (2020) propose the predictive prescription which aims to use the most similar samples in the training set to the sample to be estimated to generate the distribution, and then minimize or maximize the objective function based on the scenarios generated.

The basic idea of the predictive prescription is to assign a weight learned from data to the cost induced by the estimated values of the unknown parameters in various scenarios, so as to generate the optimal decision that minimizes the weighted-sum costs. According to Bertsimas and Kallus (2020), the weights are generated in a data-driven manner with the assistance of ML models, where k-nearest neighbors (kNN), kernel methods, local linear methods, trees, and ensembles can be used. We use the following example to demonstrate the above analysis and procedure.

2.2.1. Example 2. Maximum travel bonus problem

Suppose there is a transport network denoted by $G = (\hat{H}, A)$, where $\hat{H}$ is the set of nodes and A is the set of arcs. The travel time on arc $a \in A$ is denoted by c_a. Today, Tom is at the origin and his boss requires him to arrive at the office at the destination within time T (e.g., 60 min) to draw the monthly bonus of 100 HKD. If he is late for 1 time unit, 1 HKD will be deducted; and if he is late for 2 time units, 2 HKD will be deducted, until the bonus is reduced to 0. Similar to the solution to *Example 1*, we define $\mathbf{z}$ as a binary decision vector that represents whether an arc is traversed or not. Therefore, a non-linear integer programming model can be formulated as $\max_{z\in S}\{100 - 1 \times \min(\max(\sum_{a\in A} c_a z_a - T, 0), 100)\}$, where S is a given set that describes network constraints and $S = \{(z_1, z_2) | z_1 + z_2 = 1, z_1, z_2 \in \{0, 1\}\}$. However, in practice, the travel time on each arc c_a, $a \in A$, of the current day in unknown. Instead, we can collect the travel time c_a^i on each arc in the past N days as well as the corresponding auxiliary data $\mathbf{x}^i$ including the total distance, congestion condition, speed limit, weather, and time of day. In addition, today's auxiliary data noted by $\mathbf{x}^0$ are also known based on the weather forecast.

We first illustrate why the traditional predict-then-optimize paradigm can lead to sub-optimal decisions. For the sake of simplicity, we assume that there are only two alternative arcs a_1 and a_2 connecting the origin and the destination. Suppose that we have collected the auxiliary data in the past 2000 days which are the same as today. Suppose further that on 1000 days, the travel time on a_1 is T and that on a_2 is $T - 2$. On the other 1000 days, the travel time on a_1 is $T + 1$ and that on a_2 is $T + 2$. Given the distributions, the estimated bonus that Tom can obtain of choosing a_1 and a_2 is 99.5 and 99, respectively, which indicates that a_1 is preferred than a_2. If the distributions of a_1 and a_2 are first calculated, which should be $E(\tilde{c}_{a_1}) = T + 0.5$ and $E(\tilde{c}_{a_2}) = T$, and then input to the optimization model, path a_2 will be selected instead. Obviously, the sequential predict-then-optimize paradigm leads to a sub-optimal decision.

In the predictive prescription proposed by Bertsimas and Kallus (2020), scenarios associated with probabilities are considered. Especially, define scenario 1: $c_{a_1}^1 = T$ and $c_{a_2}^1 = T - 2$, and scenario 2: $c_{a_1}^2 = T + 1$ and $c_{a_2}^2 = T + 2$, and both scenarios have equal probability at 0.5. The optimal path is chosen by minimizing the expected bonus Tom can get considering the scenarios, i.e., $\max_{z\in S} 0.5 \times \{100 - 1 \times \min(\max(\sum_{a\in A} c_a^1 z_a - T, 0), 100)\} + 0.5 \times \{100 - 1 \times \min(\max(\sum_{a\in A} c_a^2 z_a - T, 0), 100)\}$. Consequently, it can easily be seen that a_1 will be chosen, which is in line with the optimal solution and is better than the predict-then-optimize paradigm. One last question is that we cannot have so many data as assumed in this example. To address this problem, Bertsimas and Kallus (2020) further propose to attach different weights to the decisions generated by the predicted travel time by auxiliary data considering the similarity of auxiliary data. For example, if kNN is used to find the nearest k neighbors of $\mathbf{x}^0$ which is denoted by $N_k(\mathbf{x}^0) \subset \{1, ..., N\}$, the distribution of $\tilde{\mathbf{c}}|\mathbf{x}^0$ is approximately by.

$$\Pr\left(\tilde{\mathbf{c}} = \mathbf{c}^i\right) = \begin{cases} \frac{1}{k}, i \in N_k\left(\mathbf{x}^0\right) \\ 0, i \in \{1, ..., N\} \backslash N_k\left(\mathbf{x}^0\right) \end{cases}$$

Therefore, the optimal decision is generated by solving $\max_{z\in S} \frac{1}{k} \sum_{i\in N_k(\mathbf{x}^0)} \{100 - 1 \times \min(\max(\sum_{a\in A} c_a^i z_a - T, 0), 100)\}$

3. Conclusion and future research directions

This article first discusses the importance of prediction and optimization approaches in modern decision making, especially within the era of big data. The drawbacks of the standard predict-then-optimize paradigm, which is a common way to integrate prediction with optimization tasks in a sequential way by inputting the predicted unknown parameters to the downstream optimization model, are discussed through an intuitive example. Then, two popular approaches in the existing literature to dealing with the shortcomings brought by the traditional predict-then-optimize paradigm are illustrated with examples presented, namely SPO framework and predictive prescription framework, where the decision errors caused by the inaccurate prediction of unknown parameters in the optimization model are explicitly considered and addressed in the prediction model. Research on combining and integrating prediction and optimization approaches in a unified framework is an emerging area, yet we nevertheless hope that we have provided glimpses of its potential from the existing studies covered in this article. We hereby outline some promising future research directions to enhance the integration of prediction and optimization.

First, frameworks integrating prediction with optimization that can be applied in various scenarios need further investigation. The above analysis shows that there are limitations in the applicable scenarios of each of the two popular frameworks: both of them are used to estimate unknown parameters in the objective function, while the SPO can only be applied to linear objectives and the predictive prescription is only suitable to be used when the objective is nonlinear in the unknown parameters (and it is no longer needed otherwise). In the future study, integration methods that can be applied to a wider range, e.g., in arbitrary objective function forms and for unknown parameters in constraints, should be further studied.

Second, the semi-SPO approach is worth further studying. It is widely acknowledged that the SPO framework can be computationally intractable as it might require solving the downstream optimization model thousands of times. One common solution in existing studies is to develop a surrogate of the SPO loss function to reduce the computational burden. Besides, semi-SPO, which takes the typical characteristics of the downstream optimization model into account when developing the first stage ML model while avoiding frequently solving the subsequent optimization model, is also a promising way to achieve the trade-off between high-quality decision and computation power, and thus worth more extensive investigation.

Third, the applications of the integrated frameworks of prediction and optimization to address practical transportation problems are quite limited at the moment. Therefore, more testings using either public datasets to validate and compare the performance of different frameworks or case studies to address practical transportation issues are expected in future research.

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.

Reference

Abduljabbar, R., Dia, H., Liyanage, S., Bagloee, S.A., 2019. Applications of artificial intelligence in transport: an overview. Sustainability 11 (1), 1–24.
Balghiti, O. E., Elmachtoub, A. N., Grigas, P., Tewari, A., 2019. Generalization bounds in the predict-then-optimize framework. arXiv preprint arXiv:1905.11488.
Bertsimas, D., Kallus, N., 2020. From predictive to prescriptive analytics. Manag. Sci. 66 (3), 1025–1044.
Chu, H., Zhang, W., Bai, P., Chen, Y., 2021. Data-driven optimization for last-mile delivery. Complex Intell. Syst. 2021, 1–14.
Demirović, E., Stuckey, J, Bailey, P., Chan, J., Leckie, J., Ramamohanarao, C., Guns, K., 2019. Predict+ optimise with ranking objectives: Exhaustively learning linear functions. In: Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, pp. 1078–1085 2019.
Dragomir, Voicu D., Dumitru, Mădălina, 2022. Practical solutions for circular business models in the fashion industry. Cleaner Logistics and Supply Chain 100040. doi:10.1016/j.clscn.2022.100040, In this issue.
Elmachtoub, A.N., Grigas, P., 2021. Smart "predict, then optimize. Manag. Sci. accepted and in press.
Elmachtoub, A., Liang, J.C.N., McNellis, R., 2020. Decision trees for decision-making under the predict-then-optimize framework. Int. Conf. Mach. Learn. 2020, 2858–2867.
Ferber, A., Wilder, B., Dilkina, B., Tambe, M., 2020. Mipaal: Mixed integer program as a layer. In: Proceedings of the AAAI Conference on Artificial Intelligence, 34, pp. 1504–1511.
Loke, G. G., Tang, Q., Xiao, Y., 2021. Decision-driven regularization: a blended model for predict-then-optimize. Available at SSRN 3623006.
Mandi, J., Stuckey, P.J., Guns, T., 2020. Smart predict-and-optimize for hard combinatorial optimization problems. In: Proceedings of the AAAI Conference on Artificial Intelligence, 34, pp. 1603–1610.
Peng, Bile, Keskin, Musa Furkan, Kulcsár, Balázs, Wymeersch, Henk, 2021. Connected autonomous vehicles for improving mixed traffic efficiency in unsignalized intersections with deep reinforcement learning. Communications in Transportation Research 100017. doi:10.1016/j.commtr.2021.100017, In this issue.
Schinas, Orestis, Sonechko, Dmitriy, 2022. A pay-as-you-use business model for the greening of shipping. Cleaner Logistics and Supply Chain 100034. doi:10.1016/j.clscn.2022.100034, In this issue.
Wang, Wei, Wu, Yiwei, 2021. Is uncertainty always bad for the performance of transportation systems? Communications in Transportation Research 100021. doi:10.1016/j.commtr.2021.100021, In this issue.
Wilder, B., Ewing, E., Dilkina, B., Tambe, M., 2019. End to end learning and optimization on graphs. Adv. Neural Inf. Process. Syst. 32, 4672–4683.
Yan, R., Wang, S., Fagerholt, K., 2020. A semi-"smart predict then optimize"(semi-SPO) method for efficient ship inspection. Transp. Res. Part B 142, 100–125.