

Learning for Robust Combinatorial Optimization: Algorithm and Application

Zhihui Shao
UC Riverside

Jianyi Yang
UC Riverside

Cong Shen
University of Virginia

Shaolei Ren
UC Riverside

Abstract—Learning to optimize (L2O) has recently emerged as a promising approach to solving optimization problems by exploiting the strong prediction power of neural networks and offering lower runtime complexity than conventional solvers. While L2O has been applied to various problems, a crucial yet challenging class of problems — robust combinatorial optimization in the form of minimax optimization — have largely remained under-explored. In addition to the exponentially large decision space, a key challenge for robust combinatorial optimization lies in the inner optimization problem, which is typically non-convex and entangled with outer optimization. In this paper, we study robust combinatorial optimization and propose a novel learning-based optimizer, called LRCO (Learning for Robust Combinatorial Optimization), which quickly outputs a robust solution in the presence of uncertain context. LRCO leverages a pair of learning-based optimizers — one for the minimizer and the other for the maximizer — that use their respective objective functions as losses and can be trained without the need of labels for training problem instances. To evaluate the performance of LRCO, we perform simulations for the task offloading problem in vehicular edge computing. Our results highlight that LRCO can greatly reduce the worst-case cost and improve robustness, while having a very low runtime complexity.

I. INTRODUCTION

Fast optimization with competent performance is a fundamental problem in many scientific and engineering fields. Given a problem instance, a typical optimizer often goes through multiple iterates of time-consuming gradient calculation, thus having a high computational complexity especially when the problem scales up [1]. In recent years, by leveraging the power of neural networks, learning-based optimizers (a.k.a., *learning to optimize*, or simply L2O) have quickly surfaced as efficient and effective solutions to many optimization problems, complementing or even replacing conventional optimizers [2]–[8]. The key idea of L2O is to train a neural network to learn the optimization process over a set of training problem instances and generalize to new testing problems. Compared to conventional optimizers, L2O typically has a lower computational complexity given a problem instance at runtime — L2O only needs one forward propagation over the learnt neural network. Importantly, while “labels” (i.e., solutions obtained by an existing optimizer to training problem instances) are useful, the training process of L2O can be also directly supervised by the objective function, whose gradient guides the neural network’s parameter updates. As a result, the training set does not necessarily require labels [3], [4], [6], [9].

Among the family of optimization problems, combinatorial optimization is a longstanding yet challenging problem with

numerous applications, and has received significant attention over the last few decades [10]. Notable examples range from the classic traveling salesman problem (TSP) and bin-packing problem to emerging applications such as AI model assignment and server provisioning in edge computing [11], [12]. In general, combinatorial optimization is NP-hard due to its exponentially large solution space and hence often relies on search-based methods (e.g., evolutionary algorithms [13]), approximation techniques (e.g., branch and bound [14]), and/or heuristics (e.g., greedy-based algorithms).

Recently, L2O has also been explored for combinatorial optimization, providing competent solutions with a low computation cost at runtime [9], [12], [15], [16]. For example, several studies have proposed L2O-based algorithms for various combinatorial problems, including minimum vertex cover, maximum cut, TSP, mixed-integer programming, and resource allocation in communication networks [12], [17]–[21].

Despite the recent success of L2O, a crucial class of combinatorial problems — (worst-case) robust combinatorial optimization in the form of “minimax” — have largely remained unexplored. In many practical scenarios, the input/context to the combinatorial optimization problem is uncertain or even adversarially corrupted. For example, in the wireless channel assignment problem, the channel condition for different users is a critical context but can only be estimated subject to estimation errors; in the cloud resource management problem, the user demand is the context for the resource manager, but it can only be predicted with unavoidable prediction errors; in the unit commitment problem for power plant scheduling in smart grids, the electricity demand is provided through a forecast, and hence it cannot be perfectly known a priori and can even be subject to adversarial modification [22]. In these applications, simply viewing the given context as ground truth without accounting for its imperfectness can result in bad decisions and even catastrophic consequences (e.g., power outage given an under-estimated electricity demand). Therefore, it is critically important to take into account the context uncertainty and achieve robustness in combinatorial optimization.

Nonetheless, robust combinatorial optimization is very challenging, and more so than the already-difficult combinatorial optimization alone [10]. In fact, even by relaxing the combinatorial variables and only considering continuous optimization, robust optimization through minimax is challenging, unless some additional restrictive assumptions (e.g., “convex-concave” objective function) are imposed [23]. A key reason

for the challenge is the difficulty in determining the worst case (i.e., the “max” part in minimax) that is typically non-convex. Without a reasonably good solution to the “max” part, the resulting solution may not sufficiently account for the entire uncertainty range of the context, thus failing to achieve the desired level of robustness. Moreover, the maximizer in minimax is also intrinsically entangled with the minimizer. Thus, standard L2O techniques that assume a single objective function cannot solve robust combinatorial optimization, where two competing objectives are entangled.

Contribution. In this paper, we focus on robust minimax combinatorial optimization and propose a novel learning-based optimizer, called LRCO (Learning for Robust Combinatorial Optimization), which quickly outputs a robust solution in the presence of uncertain context. As illustrated in Fig. 1, LRCO leverages a pair of learning-based optimizers — one for the minimizer and the other for the maximizer — that interact with each other during the training process and jointly find robust solutions against context uncertainty. Both the minimizer and maximizer networks use their respective objective functions as losses which, without the need of labels for training samples, can directly guide the training process. Additionally, we include an ensemble of maximizer networks to improve the performance of the maximizer.

To evaluate the performance of LRCO, we perform numerical experiments for the task offloading problem in vehicular edge computing systems. We compare LRCO with several baseline algorithms and oracle solutions. Our results highlight that LRCO can greatly reduce the worst-case utility, while also maintaining a good true utility (See Section V-C). Importantly, LRCO also has a very low runtime complexity.

II. RELATED WORKS

Learning to optimize (L2O). L2O is a general framework that mimics the optimization process by training on a set of sample problem instances. In many of the prior studies, L2O trains a recurrent neural network (RNN) to parameterize the iteration process (e.g., gradient-based updates) [2], [6]–[8], [24]. For example, concurrent studies [7], [8] employ RNN-based reinforcement learning (with known objective functions) to learn the gradient update iterations, while [25] subsequently extends the setting to unknown objective functions. In [2], a modified RNN is introduced to improve scalability and generalizability of L2O, and [26] uses LSTM models and the attention mechanism for solving Bayesian swarm optimization. To better safeguard neural networks, [5], [27] improve the adversarial training sample generation by using L2O to solve a constrained maximization problem. In the context of wireless networks, [3], [4] apply L2O to optimize power allocation in multi-user interference channels for sum rate maximization. These studies focus on continuous optimization problems and neglect the uncertainties in the input.

Combinatorial optimization. Traditionally, combinatorial optimization, such as TSP and bin-packing, relies on problem-specific heuristics/algorithms designed by domain experts [10],

[28]. More recently, L2O-based algorithms have been proposed for a variety of combinatorial problems. In general, learning-based combinatorial optimizers can use label-based supervised learning or label-free reinforcement learning [9]. In supervised learning, a large number of solutions generated based on state-of-the-art (SOTA) methods are used as labels, and the goal of the learnt optimizer is to directly mimic the solution labels or optimization steps (i.e., imitation learning) and hence reduce the computational complexity at runtime. For example, [29] proposes pointer networks for combinatorial optimization which are trained with labels from a SOTA method, and [17] uses L2O to replace the time-consuming branch and bound method for mixed-integer programming. By contrast, methods based on reinforcement learning directly learn the solution based on the reward/objective function. For example, [15] solves the TSP problem with a close-to-optimal performance, which is further improved by introducing policy gradient and attention mechanisms [30]. Learning-based combinatorial optimization over graphs (e.g., maximum cut and minimum vertex cover) is studied in [18], while [12], [19]–[21] focus on reinforcement learning-based combinatorial optimization problems in networking. In these studies, the input to the problem instance is assumed to be perfect, and uncertainty of the input is not considered.

Minimax optimization. Robust optimization is very challenging and often formulated as a minimax problem [23], [31]. The conventional solutions to continuous minimax problems are mostly based on gradient updates or problem-specific methods. For example, [32] proposes a gradient descent and gradient ascent (GDA) method, by alternatively optimizing the min function and the max function. Further, more stable GDA algorithms are also proposed, such as K-Beam [33], optimistic-GDA [34], [35] and Follow-the-Ridge [36].

A generative adversarial network also involves minimax optimization [37], but it typically requires true samples/labels to train the discriminator for distinguishing fake samples from true ones. By contrast, LRCO learns two optimizers (min and max), and the notion of “true” samples does not apply.

Because of its combinatorial nature, robust combinatorial optimization has remained relatively under-explored, with some non-learning-based algorithms available under specific settings and assumptions (e.g., [31], [38] consider monotone and submodular functions with a robust setting where a certain number of actions can be nullified).

The very recent study [23] uses L2O to solve *continuous* minimax problems, but our work is substantially different. Concretely, [23] trains two individual LSTM networks using a self-defined loss function, and uses these two networks to replace the gradient update process for both the min and max parts (used in conventional optimizers like [32]). Thus, this approach does not apply to robust combinatorial optimization which, unlike continuous optimization, does not have standard gradient updates and is more challenging. Additionally, unlike [23] using a single network for the max part, we include an ensemble of networks to further improve the maximizer performance.

In summary, LRCO advances the quickly expanding field of L2O and is the first learning-based optimizer offering a fast and better solution to *robust* combinatorial optimization.

III. PROBLEM FORMULATION

Combinatorial optimization problems widely exist in scientific and engineering applications, including resource allocation, scheduling, and capacity planning in computing and networking systems [12]. In general, a combinatorial optimization problem can be written as $\min_{\mathbf{a} \in \mathcal{A}} f(\mathbf{x}, \mathbf{a})$, where \mathbf{a} is the decision variable, \mathcal{A} is the feasible decision set, and $f(\cdot, \cdot)$ indicates the objective function with the context parameter \mathbf{x} provided to the optimizer. For generality, we note that both the context \mathbf{x} and decision \mathbf{a} are *vectors* but with possibly different dimensions. This formulation applies to integer programming, which also belongs to combinatorial optimization.

A key novelty that distinguishes our work apart from standard combinatorial optimization is consideration of the uncertainty in the context parameter \mathbf{x} , as motivated by practical applications. For example, before deciding the optimal channel assignment for users in a wireless network, the channel condition (i.e., context \mathbf{x} in our formulation) is needed, but it can only be estimated. To capture the context uncertainty, we assume that the true context is $\mathbf{x} + \delta$, where the context error δ belongs to an uncertainty set Δ (also referred to as uncertainty/error budget in robust optimization). In the robust optimization and learning literature [23], [39], δ is commonly bounded by a L_p norm, i.e., $\Delta = \{\delta : |\delta|_p \leq \epsilon\}$ where $|\delta|_p$ denotes the L_p norm of δ with $p \geq 1$. While we consider continuous context parameters \mathbf{x} for the ease of presentation, LRCO can also be extended to discrete contexts by modifying the learning-based maximizer (Section IV-B) such that it solves combinatorial optimization.

We now formulate the robust combinatorial optimization problem as follows:

$$\min_{\mathbf{a} \in \mathcal{A}} \max_{\delta \in \Delta} f(\mathbf{x} + \delta, \mathbf{a}) \quad (1)$$

where \mathbf{x} is the uncertain context available to the optimizer and $\delta \in \Delta$ denotes the context error unknown to the optimizer. Our minimax formulation captures the *worst-case* robustness over the entire uncertainty set Δ , and is commonly used as a robust metric [23], [31], [38]. Note that the actually achieved cost may not be the worst case and hence can often be lower than worst-case cost.

Challenges. Even with the perfect context (i.e., $\delta = 0$), combinatorial optimization is in general NP-hard and can only be solved by approximate algorithms [10], [12], [28]. Furthermore, the uncertain context $\delta \in \Delta$ makes robust combinatorial optimization even more challenging, because it is time-consuming to use a conventional solver to solve the inner maximization problem in Eqn. (1) due to its non-convexity in many practical applications. Thus, except for special cases, it is generally impossible to have a closed-form solution for the inner problem $\max_{\delta \in \Delta} f(\mathbf{x} + \delta, \mathbf{a})$. As a result, the outer minimization problem cannot be efficiently solved

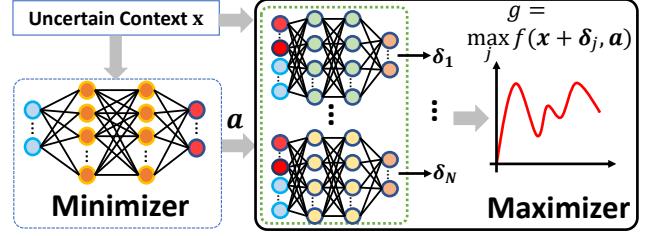


Fig. 1. Overview of LRCO. The maximizer network produces a set of candidate actions, which are then evaluated by the minimizer for robustness. The action with the minimum worst-case cost is selected.

either. Even though we can use a learning-based minimizer [12], [18] along with a conventional maximizer, the runtime computational complexity given a new problem instance is very high, because the learning-based minimizer produces a probability distribution for decisions and the conventional maximizer needs to be executed once for each sampled decision in order to select the optimal robust decision. As a preview of the numerical results (Section V), the solver is about 70x slower than our learning-based maximizer, and the maximizer needs to find the worst-case cost for each of the 1000 candidate decisions sampled by the minimizer. On the other hand, while simple heuristic methods (e.g., greedy) can be fast, they may not offer a satisfactory performance. To address these challenges, we shall present LRCO, a learning-based optimizer that produces a competent solution \mathbf{a} with a low computational complexity at runtime.

IV. LRCO: LEARNING FOR ROBUST COMBINATORIAL OPTIMIZATION

In this section, we provide the design of LRCO, which combines two learning-based optimizers (i.e., min and max) for efficiently solving robust optimization problems.

A. Overview

The key idea of LRCO is to use machine learning models, in particular neural networks, to parameterize the minimizer and maximizer for improving computational efficiency given new problem instances at the runtime. We choose neural networks mainly because of their strong universal approximation capability [40]. An overview of LRCO is shown in Fig. 1, which includes two learning-based optimizers — a minimizer and a maximizer that solve the outer and inner part of Eqn. (1), respectively.

Testing/inference. Given a problem instance with uncertain context at runtime, the optimization process is described in Algorithm 1. First, the neural network-based minimizer outputs the probability of decisions, based on which a number of candidate decisions are sampled accordingly. For the maximizer which typically involves non-convex optimization, we use an ensemble of neural networks. Specifically, given each pair of uncertain context \mathbf{x} and sampled decision \mathbf{a} (produced by the minimizer), the maximizer obtains the worst-case context error δ through the ensemble. Finally, the decision that has the minimum worst-case cost is chosen. At runtime,

Algorithm 1 LRCO for Testing/Inference

- 1: **Inputs:** Uncertain context \mathbf{x} , minimizer network parameter θ_a , and ensemble maximizer network parameters $\theta_{w,i}$ for $i = 1, \dots, N$
- 2: **Output:** Robust decision \mathbf{a}
- 3: Use the minimizer network to generate action probability distribution based on Eqn. (3), and sample K candidate decisions \mathbf{a}_k accordingly
- 4: **for** $k = 1, \dots, K$ **do**
- 5: **for** $i = 1, \dots, N$ **do**
- 6: Use the maximizer network i to generate the worst-case context error $\delta_{i,k}$ for decision \mathbf{a}_k
- 7: Calculate $g_{i,k} = f(\mathbf{x} + \delta_{i,k}, \mathbf{a}_k)$
- 8: **end for**
- 9: Find the worst-case cost $G_k = \max_i g_{i,k}$ for \mathbf{a}_k
- 10: **end for**
- 11: Select the decision $\mathbf{a} = \mathbf{a}_k$, where $k = \arg \min_k G_k$

the entire optimization process in LRCO only involves forward propagation over the minimizer and maximizer neural networks, and hence has a much lower complexity than using conventional solvers.

Training. For training the neural networks, we do not provide ground-truth sample solutions as labels, i.e., solutions obtained by an existing solver. One reason is that it can be very time-consuming to generate solutions to many problem instances (even offline) due to the combinatorial nature in the minimizer and non-concavity in the maximizer. In addition, for complex optimization problems, solutions directly learnt by neural networks may even exceed those obtained by conventional solvers [4]. Thus, we use the objective function $f(\cdot, \cdot)$ as the loss to directly supervise the training. Note that, like other machine learning methods, the goal of LRCO is to speed up the runtime inference for testing samples (compared to non-learning methods), although it requires additional offline training time.

LRCO vs. reinforcement learning. The architecture of LRCO resembles the actor-critic reinforcement learning [41]: in LRCO, the minimizer's decision is evaluated by the maximizer, while the actor's policy is evaluated by the critic in reinforcement learning. Nonetheless, there are also crucial differences. First, in LRCO, the training is performed entirely offline based on the distribution of context \mathbf{x} since the objective function $f(\cdot, \cdot)$ is already known, whereas reinforcement learning is often updated online to adapt to a new environment. Second, the actor in reinforcement learning typically learns the value function based on reward signal feedback (i.e., labels), whereas the maximizer in LRCO leverages L2O to solve an (often non-convex) optimization problem.

B. Maximizer Network

We now present the details of our learning-based maximizer, which includes an ensemble of neural networks to efficiently produce a solution to the inner maximization problem in $\min_{\mathbf{a} \in \mathcal{A}} \max_{\delta \in \Delta} f(\mathbf{x} + \delta, \mathbf{a})$.

Algorithm 2 Iterative training of LRCO

- 1: **Inputs:** Training set of uncertain context \mathcal{D}_x ,
- 2: **Output:** Minimizer network parameter θ_a , and ensemble maximizer network parameters $\theta_{w,i}$ for $i = 1, \dots, N$
- 3: **Initialize:** Randomly generate a training set of decisions $\mathcal{D}_{\mathbf{a}}^0$, and pre-train the ensemble of maximizer networks $\theta_{w,i}$ for $i = 1, \dots, N$ over \mathcal{D}_x and $\mathcal{D}_{\mathbf{a}}^0$
- 4: **for** $k = 1, \dots, \text{MaxIterate}$ **do**
- 5: Train the minimizer network using Algorithm 3
- 6: **for** $\mathbf{x} \in \mathcal{D}_x$ **do**
- 7: Calculate $P_{\theta_a}(\mathbf{a}|\mathbf{x})$ based on Eqn. (3)
- 8: **end for**
- 9: Calculate the new distribution of decisions $P_{\theta_a}(\mathbf{a}) = \frac{1}{|\mathcal{D}_x|} \sum \mathbf{x} \in \mathcal{D}_x P_{\theta_a}(\mathbf{a}|\mathbf{x})$
- 10: Randomly generate a new training set of decisions $\mathcal{D}_{\mathbf{a}}^k$ based on the new distribution $P_{\theta_a}(\mathbf{a})$
- 11: Train the ensemble of maximizer networks $\theta_{w,i}$, for $i = 1, \dots, N$, over the dataset \mathcal{D}_x and $\mathcal{D}_{\mathbf{a}}^k$
- 12: **end for**

In general, the inner maximization problem is non-convex (e.g., the vehicular task offloading problem in Section V). Alternatively, one may want to use a conventional solver (e.g., based on gradient updates [1]) for the inner maximization problem, but this has a high computational cost. Specifically, for testing at runtime, the minimizer produces a probability distribution for candidate decisions, and the maximizer needs to evaluate the worst-case cost for each sampled decision (e.g., 1000 in total in Section V) in order to select the optimal decision. Thus, using conventional solvers can incur a very high computational complexity.

In LRCO, we train the maximizer network by directly solving the inner optimization problem $\max_{\delta \in \Delta} f(\mathbf{x} + \delta, \mathbf{a})$ where $\Delta = \{\delta, |\delta|_p \leq \epsilon\}$ is the uncertainty set. Specifically, by viewing the uncertain context \mathbf{x} and (candidate) decision \mathbf{a} as the input, the maximizer network learns to optimize δ , as supervised by the training loss function below:

$$\mathcal{L}(\mathbf{x}, \mathbf{a}, \delta) = -f(\mathbf{x} + \delta, \mathbf{a}) + \lambda [|\delta|_p - \epsilon]^+, \quad (2)$$

where we add the minus sign to be consistent with the definition of loss function, and $\lambda [|\delta|_p - \epsilon]^+$ means that additional penalty weighted by $\lambda > 0$ is imposed whenever the solution δ goes beyond the uncertainty set Δ . As the neural network is trained based on sampled problem instances, the maximizer network can ensure by tuning λ that the $[|\delta|_p - \epsilon]^+$ term is sufficiently small on average. In case that the output δ violates the uncertainty set Δ for a particular testing sample, we can scale down δ to force it to fall into Δ .

To train the maximizer network, we first generate samples based on the distribution of uncertain context \mathbf{x} and candidate decisions \mathbf{a} . Then, with training samples and the loss function in Eqn (2), any standard learning approaches, like stochastic gradient descent, can be applied, and hence we omit the description. Since the inner optimization problem

$\max_{\delta \in \Delta} f(\mathbf{x} + \delta, \mathbf{a})$ is typically non-convex, simply using one maximizer network may not produce a satisfactory solution. Thus, for the maximizer, we use an ensemble of neural networks, each initialized with a different weight. Prior studies [4] have shown that the ensemble approach can significantly improve the optimization performance.

We note that the distribution of decisions \mathbf{a} is entangled with the minimizer network's output. One approach is to randomly generate candidate decisions covering a sufficiently wide distribution. While the resulting maximizer network can produce good solutions to a wide range of decision distributions, the actual distribution of the decisions produced by the minimizer network is only a subset of the distribution used for training. In other words, the capacity of the maximizer neural networks is not fully utilized. Here, we first randomly generate \mathbf{a} and pre-train the minimizer network, use it to evaluate the worst-case to guide the minimizer network's training, and then update the training set of \mathbf{a} based on the new distribution of decisions produced by the minimizer network. This process can repeat a few iterations as shown in Algorithm 2. While we observe convergence in all our empirical experiments, it is an interesting future study to derive rigorous conditions under which the iterative training process is guaranteed to converge.

C. Minimizer Network

The learning-based minimizer employs a neural network that parameterizes the solution to the outer combinatorial optimization problem in $\min_{\mathbf{a} \in \mathcal{A}} \max_{\delta \in \Delta} f(\mathbf{x} + \delta, \mathbf{a})$. Instead of generating labels (i.e., solutions obtained by an existing solver) for supervision, we use a reinforcement learning process to train the minimizer network. Although there exist many reinforcement learning-based combinatorial optimizers [9], [12], [15], [16], they are directly supervised by the objective function under the assumption of perfect context parameters. In sharp contrast, to account for uncertain context parameters, the minimizer network in LRCO uses the worst-case cost returned by our learning-based maximizer (Section IV-B) for supervision. Additionally, due to the coupling of minimization and maximization in $\min_{\mathbf{a} \in \mathcal{A}} \max_{\delta \in \Delta} f(\mathbf{x} + \delta, \mathbf{a})$, the output of our minimizer network also influences the training of the maximizer (see Algorithm 2).

Concretely, the minimizer network parameterized by θ_a takes the uncertain context \mathbf{x} as input and yields the probability of candidate decisions $p_{\theta_a}(\mathbf{a}|\mathbf{x})$ over the feasible decision set \mathcal{A} . The parameter θ_a is trained by following a policy gradient algorithm under the supervision of our minimizer (which plays a similar role of the “critic” in reinforcement learning) [41]. The training process is described in Algorithm 3.

Decision representation. A well-trained minimizer network should assign higher probabilities to better decisions that are more likely to minimize the worst-case cost. For combinatorial optimization, the feasible decision set is typically exponentially large. For example, a 10-dimensional decision \mathbf{a} , with 5 possible integer values for each dimension, will result in 5^{10} decisions. Thus, it is difficult to directly represent all

Algorithm 3 Minimizer Network Training

```

1: Inputs: Training set of uncertain context  $\mathcal{D}_x$ , and ensemble network for the maximizer, training epochs  $T$ , batch size  $B$ , sampling size  $|\mathcal{S}|$ 
2: Output: Minimizer network weight parameter  $\theta_a$ 
3: Initialize: Randomly initialize the parameter  $\theta_a$ 
4: for  $t = 1, \dots, T$  do
5:   for  $i = 1, \dots, B$  do
6:     Randomly sample  $\mathbf{x}$  from  $\mathcal{D}_x$ 
7:     Calculate decision probability  $P_{\theta_a}(\mathbf{a}|\mathbf{x})$  based on Eqn. (3), and sample decisions from  $P_{\theta_a}(\mathbf{a}|\mathbf{x})$ 
8:     Run the maximizer (Section IV-B) to find the worst-case cost  $G = \max_{\delta \in \Delta} f(\mathbf{x} + \delta, \mathbf{a})$ 
9:     Generate a random set of decisions  $\mathcal{S}$  from  $\mathcal{A}$ 
10:    for  $\mathbf{d}_s \in \mathcal{S}$  do
11:      Run the maximizer to find the worst-case cost  $G_s = \max_{\delta \in \Delta} f(\mathbf{x} + \delta, \mathbf{a}_s)$ 
12:    end for
13:    Calculate the baseline  $V(\mathbf{x}) = \frac{1}{|\mathcal{S}|} \sum_{s=1}^{|\mathcal{S}|} G_s$ 
14:    Calculate gradient  $-[G - V(\mathbf{x})] \nabla_{\theta_a} P_{\theta_a}(\mathbf{a}|\mathbf{x})$ 
15:  end for
16:  Update  $\theta_a$  with batched gradient
17: end for

```

the decisions using the conventional one-hot encoding, where each node in the output layer represents a single possible decision. To address this issue, we divide the decision \mathbf{a} into D decision groups. Within each group i , we use an integer a_i to represent the feasible decisions using one-hot encoding in the final output layer, followed by softmax activation which produces the probability distribution for each decision group.¹ Then, by rewriting the decision vector as $\mathbf{a} = [a_1, \dots, a_D]^T$, we factorize the final decision probability as follows:

$$P_{\theta_a}(\mathbf{a}|\mathbf{x}) = \prod_{i=1}^D p_{\theta_a}(a_i|\mathbf{x}). \quad (3)$$

To further illustrate this point, consider that we have 10 communication channels and need to decide whether to occupy each channel. For this case, we have 10 independent decision groups, each with a binary decision. If it is not possible to divide the decision \mathbf{a} into independent decision groups, we can temporarily ignore the coupling constraint and then re-scale the decision probability. For example, using the previous example and assuming a cardinality constraint B on the set of selected channels, we can obtain $P_{\theta_a}(\mathbf{a}|\mathbf{x}) = \prod_{i=1}^D p_{\theta_a}(a_i|\mathbf{x})$ and only consider those decisions with the B highest probabilities. While we present a general approach, we can also exploit applicable problem structures, e.g., graph topology, to better encode the decisions (see [18] for an example).

Policy gradient. As in the Monte-Carlo reinforced policy gradient approach [15], [41], the net cost benefit of a decision $\max_{\delta \in \Delta} f(\mathbf{x} + \delta, \mathbf{a}) - V(\mathbf{x})$ is used for gradient calculation,

¹If a decision group only contains one binary decision, we can then use only one output node followed by sigmoid activation for that decision group.

instead of the absolute cost $\max_{\delta \in \Delta} f(\mathbf{x} + \delta, \mathbf{a})$. Here, $V(\mathbf{x})$ is the baseline value function calculated based on a set of randomly sampled decisions (Line 13 of Algorithm 3). This can effectively avoid unnecessarily penalizing those decisions that are good (relative to the baseline) but have low absolute values. The sampling process in Line 7 of of Algorithm 3 also encourages exploration without always exploitation.

We can also employ an ensemble of neural networks for the minimizer, although our results in Section V shows that a single neural network has already performed very well.

V. APPLICATION: TASK OFFLOADING IN VEHICULAR EDGE COMPUTING

To evaluate LRCO in real networking applications, we apply LRCO to the problem of replicated task offloading in a vehicular edge computing (VEC) system. We first describe the problem formulation, then explain the simulation setup, and finally present the results under two different error budgets.

A. Background and Problem Formulation

With the advances in connected driving and vehicular networks, VEC is emerging as a promising computing architecture, complementing the conventional cloud systems. Specifically, to improve the service quality and user experience, computation is gradually moving toward the edge of vehicular networks, and vehicles with extra computation resources can even provide distributed computing to other vehicles and nodes. We refer to those vehicles that can provide computation on the move as *vehicular clouds*. Nonetheless, because of the continuously changing environment (e.g., vehicle's locations and/or server resource availability), the resource management decisions in a VEC system must also be highly agile and adaptive — the resource manager needs to dynamically and efficiently assign the workloads to different servers. Furthermore, performance robustness is crucial in VEC, especially for safety-critical applications. Next, we consider the replicated task offloading problem in VEC and formulate it into robust combinatorial optimization.

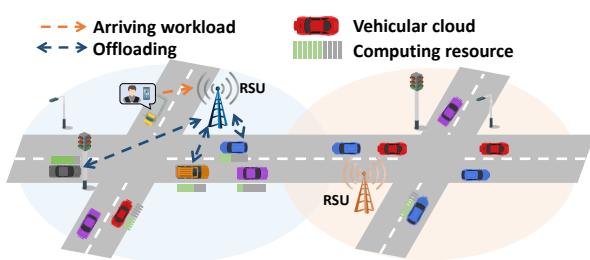


Fig. 2. Illustration of task offloading in a VEC system.

Consider a VEC system as illustrated in Fig. 2, including a road-side-unit (RSU) and multiple vehicular clouds that can provide computation on demand. The RSU is installed at a fixed location, which gathers computing demand/tasks from some vehicles and then offloads them to nearby available vehicular clouds via wireless communications. Each workload

includes multiple micro services, which are offloaded to different vehicular clouds. The service delay d for each micro service includes the data transmission delay d^t of the wireless network and service computing delay d^c in the respective vehicular cloud. By default, each task arrives with a maximum delay constraint L . The task is successfully finished if all involved micro services are completed prior to the deadline, and fails otherwise.

Due to the highly dynamic environment (e.g., fast-changing locations of vehicular clouds), it is difficult to guarantee the success of each offloaded task, and estimating the success rate of each offloading decision a priori is also non-trivial. Thus, to achieve a higher success rate, each arriving task is replicated and offloaded to multiple vehicular clouds as in [42], [43]. In addition, it is important to design a robust offloading policy for the RSU, improving the overall success rate even in the worst case. We denote x_{ij} as the success rate when a micro service j is offloaded to vehicular cloud v_i . Then, the overall task success rate can be calculated via the *At-Least-One* rule as shown below:

$$P(\mathbf{x}, \mathbf{a}) = \prod_{j=1}^M \left[1 - \prod_{i=1}^C (1 - x_{ij} a_{ij}) \right] \quad (4)$$

where $a_{ij} \in \{0, 1\}$ represents the offloading decision (i.e., “1” means offloading the micro service j to vehicular cloud i , and “0” means otherwise), M is the total number of micro services in the task, and C is the maximum number of vehicular clouds to which each micro service can be offloaded. The interpretation of Eqn. (4) is that each micro service should be successfully completed by at least one vehicular cloud, and the overall task is successful only when all the micro services are successfully executed.

Besides the success rate, there is also a cost of each offloading (e.g., the computational cost incurred, incentives or payment made to the vehicular cloud). Here, we denote η_{ij} as the cost for offloading micro service j to vehicular cloud i . Thus, the expected utility/reward given the offloading decision can be formulated as

$$U(\mathbf{x}, \mathbf{a}) = \prod_{j=1}^M \left[1 - \prod_{i=1}^C (1 - x_{ij} a_{ij}) \right] - \sum_{j=1}^M \sum_{i=1}^C \eta_{ij} a_{ij} \quad (5)$$

where x_{ij} is the context parameter, and a_{ij} is the offloading decision, for $i \in \{1, 2, \dots, C\}$ and $j \in \{1, 2, \dots, M\}$.

In Eqn. (5), the success rate (i.e., context parameter) x_{ij} for each replication is usually predicted by an estimation model based on the environment conditions. We denote the feature as c_{ij} , which includes the distance l from RSU to the vehicular cloud, the CPU utilization cpu of the vehicular cloud, and task's deadline. Naturally, no matter how accurately x_{ij} is estimated based on the feature c_{ij} , there exist estimation errors. Simply ignoring the uncertainty in x_{ij} can result in arbitrarily bad consequences. Thus, to provide a level of assurance, we need to consider robust optimization formulated as follows:

$$\max_{\mathbf{a} \in \mathcal{A}} \min_{\delta \in \Delta} U(\mathbf{x} + \delta, \mathbf{a}), \quad (6)$$

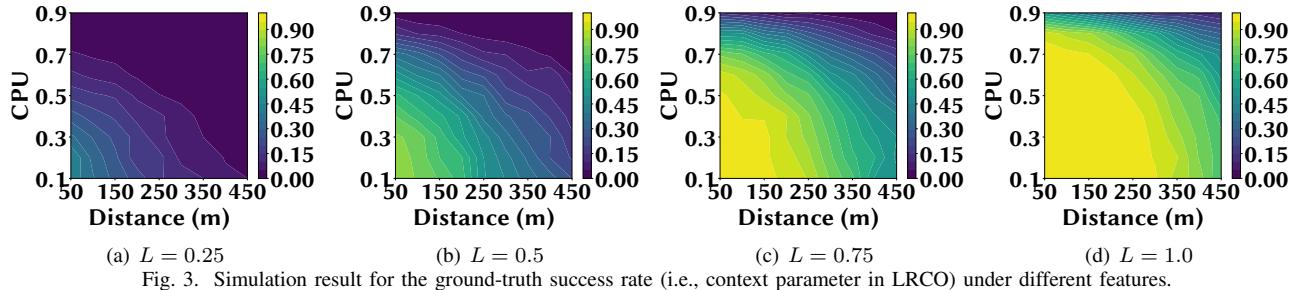


Fig. 3. Simulation result for the ground-truth success rate (i.e., context parameter in LRCO) under different features.

where we consider an L_2 -norm uncertainty set $\Delta = \{\delta, \|\delta\|_2 \leq \epsilon\}$. While the greedy algorithm has a performance guarantee in terms of competitive ratio due to the submodularity of the objective function when the context parameter \mathbf{x} is perfect [28], our robust combinatorial optimization problem in Eqn. (6) is much more challenging and the greedy algorithm can have a poor performance (see Section V-D). Finally, when using LRCO, we add a minus sign in the utility function in Eqn. (6) to convert the maximin problem into the standard minimax form.

B. Simulation Setup

We now describe how we estimate the context parameter, calculate the *ground-truth* context in our simulation, and prepare the dataset.

Context parameter estimation. We estimate the context parameter based on the respective feature information. In a real-world application, the selection of prediction models is determined by the trade-off between interpretability and accuracy. For example, a simple linear model is explainable but often has a higher prediction error, whereas a neural network model can predict with a lower error, but lacks good interpretability. In our study, we consider two prediction models: empirical linear model and residual model implemented with a neural network. More details about prediction models are presented in the appendix. The linear model captures the high uncertainty budget case (i.e., larger ϵ), whereas the neural network residual model corresponds to a low uncertainty budget.

Latency model. A micro service offloaded to v_i is successful when the computation finishes within the deadline L , i.e., $d = d^t + d^c \leq L$. The transmission time d^t can be calculated based on the wireless channel model as follows:

$$d^t = \frac{S_{data}}{W \cdot \log_2 \left(1 + \frac{P \cdot (d)^{-\alpha}}{\sigma^2 + I^t} \right)} \quad (7)$$

where S_{data} is the offloading data size, P is the transmission power, α is the channel gain factor with respect to the RSU-vehicle distance, σ^2 is the random noise, and I^t indicates the interference noise. We set the parameter as followings: $S_{data} = 3Mb$, $W = 10M$, $P = 10dBm$, $\sigma^2 = -172dBm$, $I^t \in [-10, -30]dBm$ and $\alpha = 1.8$ in our simulation. Besides, the measurement error for the vehicle's location is set as 3% (10m) according to the recent GPS manual [44]. Then, the CPU-related computation time is simulated and obtained

based on the M/M/1 queueing process [45] using the dataset in [46]. Concretely, our computation latency model follows $d^c = \frac{a}{b \cdot \text{cpu}} + \text{noise} \simeq \frac{0.227}{2.15 - \text{cpu}} + 0.007 \cdot \mathcal{N}(0, 1)$, where cpu is the CPU utilization.

Dataset preparation. By utilizing the latency model described above, we can generate the ground-truth context parameter dataset with the resulting success rate. The success rate for each offloading decision is calculated as the average over 1000 simulation rounds. The results are illustrated in Fig. 3 with different features. In total, by considering random features (i.e., distance, CPU, and deadline), we generate a training dataset with 15k instances, a validation dataset with 4k instances, and a test dataset with 6k instances.

Each task includes $M = 4$ micro services, each offloaded to up to $C = 5$ vehicle clouds, resulting a complexity of $\mathcal{O}(2^{20})$. Thus, each problem instance in our dataset contains a 20-dimensional context vector. Based on the context prediction models in the appendix, we investigate two error budgets for robustness — 0.71 for the empirical model, and 0.27 for the neural network model.

C. Baseline Algorithms, Performance Metrics, and Training

Baselines. We consider the following three baseline algorithms and two oracles for comparison.

- **Random:** The offloading decisions are randomly made without considering the context parameter or uncertainties.
- **Greedy:** It greedily solves the problem, starting from zero decisions (i.e. $a_{ij} = 0$) and based on the predicted context.
- **LCO (Learning for Combinatorial Optimization):** It is similar to LRCO, but oblivious of context uncertainty and directly uses the predicted utility (i.e., $U(\mathbf{x}, \mathbf{a})$) to train a neural network for optimization.
- **Weak Oracle:** It uses exhaustive search to maximize the predicted utility $U(\mathbf{x}, \mathbf{a})$ without considering uncertainty.
- **Oracle:** It knows the true context and uses exhaustive search to directly maximize the true utility.

Note that due to non-convexity of the inner optimization problem of Eqn. (6), constructing another oracle that exhaustively searches for the optimal robust decision (e.g., using a SOTA solver to solve the inner problem for each sampled candidate decision) is beyond our computational capability.

Performance Metrics. We consider the following metrics.

- **Predicted utility:** It is the utility by viewing the predicted context \mathbf{x} as the true one.
- **True utility:** It is the utility under the true context.

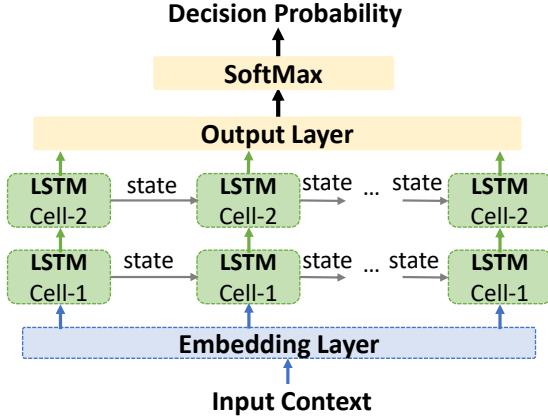


Fig. 4. Architecture of the minimizer network in our simulation.

- Worst-case utility: It is the worst-case utility over the entire the context uncertainty set.

The worst-case context may not be the true one. Thus, in general, there exists a tradeoff between worst-case utility and the true utility — more emphasis on the worst-case robustness performance can compromise the average performance. On the other hand, the predicted utility has no practical meaning. In this paper, we focus on robustness, while also showing the predicted and true utility for reference.

During the testing phase, to avoid biases of using our own maximizer when calculating the worst-case utility (i.e., our maximizer may favor LRCO), we use the solver in *scipy.optimize* package to replace our maximizer and solve the inner optimization problem. The decision in LRCO is still made by using our own maximizer. In other words, given a decision \mathbf{a} (by any baseline/oracle algorithm, or LRCO), we use the solver to calculate $\min_{\delta \in \Delta} U(\mathbf{x} + \delta, \mathbf{a})$. Note that the solver is much slower (i.e., $\sim 70x$ in our experiment) than running forward propagation in our ensemble of maximizer networks.

Training. The minimizer network is implemented with a neural network with an embedding layer, two LSTM layers with cell-1 with 50 hidden nodes and cell-2 with one hidden node, and a fully-connected output layer with $N = 20$ nodes and softmax activation. The detailed architecture is illustrated in Fig. 4. The model is trained by Adam optimizer [47]. To avoid gradient explosion, an exponentially decreasing learning rate and gradient capping are employed. More specifically, the learning rate is initialized as $1e^{-3}$ and decreases by a factor of 0.9 for every 20 epochs. In the inference/testing phase, by default, 1000 decisions are sampled based on the output probability $P_{\theta_a}(\mathbf{a}|\mathbf{x})$, and an optimal decision is selected by the maximizer. Note that LCO is also implemented with the same network architecture, but trained by directly using the predicted utility $U(\mathbf{x}, \mathbf{a})$.

We train an ensemble of four neural networks to solve the inner worst-case part in (6). According to Eqn. (6), the performance is only impacts by $\delta_{i,j}$ and $\mathbf{a}_{i,j}$. Thus, each network includes two hidden layers, each with 400 neurons, and one

TABLE I
AVERAGE UTILITY UNDER A LARGE ERROR BUDGET.

Algorithm	Predicted	True	Worst Case
Random	0.3408	0.1398	-0.0694
Greedy	0.7395	0.4103	0.1075
LCO	0.7545	0.4289	0.1206
Weak Oracle	0.7812	0.4441	0.1234
Oracle	0.7055	0.6345	0.1600
LRCO	0.6481	0.5357	0.3940

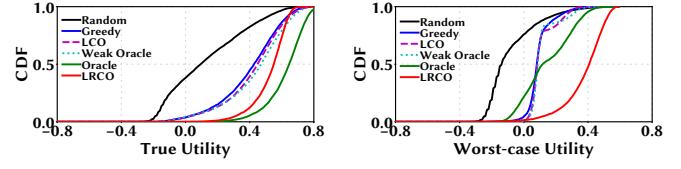


Fig. 5. CDF plots of utilities under a large error budget. (a) True utility. (b) The worst-case utility.

customized output layer, whose node value is multiplied by $\mathbf{a}_{i,j}$ (such that $\delta_{ij} = 0$ when $\mathbf{a}_{i,j} = 0$). These networks are trained using the same dataset, but different initial weights and weight parameter λ in the loss function in Eqn. (2).

D. Results with A Large Error Budget

We now investigate the performance of LRCO with a large context parameter error budget $\Delta = \{\delta, |\delta|_2 \leq 0.71\}$, when an empirical linear model is used for success rate prediction.

We show the average performances of different algorithms on the test dataset summarized in Table I. The worst-case utility is our focus, while the true utility reflects how well an algorithm performs on average in the typical cases. The results show that LCRO provides the best performance among all algorithms (including the oracle methods) under the worst-case condition, demonstrating its strong robustness. Meanwhile, the true utility of LCRO outperforms the other algorithms, except for Oracle that knows the true context in advance. Weak Oracle has the highest predicted utility because it directly maximizes the predicted utility, but the predicted utility has no practical meaning. Importantly, by comparing LRCO with LCO, we see that being oblivious of the context uncertainty can take a high toll in terms of the worst-case utility and robustness. Moreover, LCO can provide near-optimal predicted utility compared with Weak Oracle, confirming that the learning-based optimizer can be used to solve combinatorial optimization.

Additionally, we visualize in Fig. 5 the cumulative distribution function (CDF) of true utilities and worst-case utilities for different algorithms. These results show that LRCO can provide near-optimal true utility performance, although there is a gap due to the consideration of robustness in the presence of a large error budget. Further, LCRO can provide better robust performance than oracle algorithms.

E. Results with A Small Error Budget

Now, we turn to the case of a small error budget $\Delta = \{\delta, |\delta|_2 \leq 0.27\}$, by using a neural network context predictor

TABLE II
AVERAGE UTILITY UNDER A SMALL ERROR BUDGET.

Algorithm	Predicted	True	Worst Case
Random	0.1564	0.1398	-0.0022
Greedy	0.6150	0.5760	0.3763
LCO	0.6421	0.5988	0.3988
Weak Oracle	0.6553	0.6117	0.4165
Oracle	0.6331	0.6345	0.4176
LRCO	0.5988	0.5810	0.4597

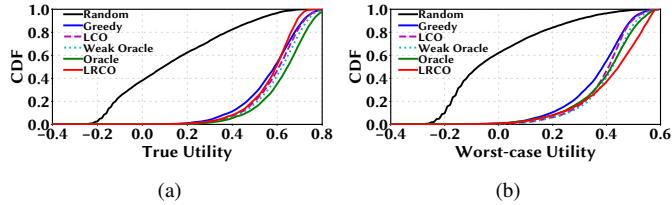


Fig. 6. CDF plots of utilities under a small error budget. (a) True utility. (b) The worst-case utility.

TABLE III
ENSEMBLE SETTINGS FOR LCRO.

Setting	Predicted	True	Worst Case
Ensemble (default)	0.6481	0.5357	0.3940
w/o Ensemble ($\lambda = 1$)	0.6423	0.5289	0.3842 ↓
w/o Ensemble ($\lambda = 10$)	0.6396	0.5274	0.3830 ↓

described in Appendix. All other settings remain unchanged.

We show in Table II the average utility under different algorithms and metrics. Also, the utility CDF is shown in Fig. 6. In the case of a small context prediction error, we have similar observations as in the case of a large prediction error in Section V-D. As intuitively expected, the gap between the true utility and the worst-case utility becomes smaller, since the context uncertainty set is smaller. Importantly, among all the algorithms, LRCO still has the highest worst-case utility and robustness.

F. Sensitivity Study

We perform the sensitivity study for $\Delta = \{\delta, |\delta|_2 \leq 0.71\}$, while the results for $\Delta = \{\delta, |\delta|_2 \leq 0.27\}$ are also similar and hence omitted due to space limitation.

Impact of ensemble settings. In LCRO, the maximizer is comprised of an ensemble of four neural networks, considering the non-convex nature of $U(\mathbf{x}, \mathbf{a})$. Here, we present the comparison results with and without ensembling in Table III, including two λ settings. We see that the ensemble approach provides a better and more robust performance, although having a single neural network for the maximizer only has minor degradation.

Impact of minimizer network size. We vary the hidden nodes to 20 and 200. The results are shown in Table IV, demonstrating that a larger network can provide a slightly better performance although the training and inference time can also increase.

Impact of sampling count. In the testing phase, by default, the best decision is selected by sampling 1000 candidate decisions based on the probability produced by the minimizer

TABLE IV
MINIMIZER NETWORK SIZE FOR LCRO.

# of Hidden Nodes	Predicted	True	Worst Case
20	0.6436	0.5299	0.3888 ↓
50 (default)	0.6481	0.5357	0.3940
200	0.6480	0.5358	0.3941

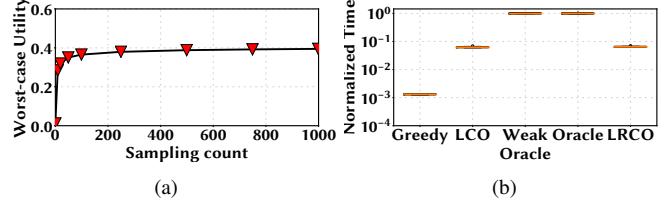


Fig. 7. (a) Performance with different sampling counts. (b) Inference time complexity for different algorithms.

network. Now, we show the performance with different numbers of samples. The robust performance is shown in Fig. 7(a). In general, the more samples, the better performance. This comes at a cost of increasing the inference time, as more candidate decisions need to be evaluated for the worst-case utility by our maximizer in LRCO.

G. Inference Time

A key advantage of LRCO compared with the existing solvers is the fast inference time at runtime, although it incurs extra training cost offline. The time is recorded with ‘time’ package in Python3. We run 10 times over all the testing data for each algorithm. Finally, the average recorded time over 10 runs is normalized with respect to the time of Oracle. The normalized time is illustrated with box-plot in Fig. 7(b). The results demonstrate that LRCO can provide compelling solutions much faster ($\sim 10x$) than oracle algorithms. While Greedy is fast and can provide a good performance for the predicted utility, it is not robust. Also, we see that LRCO is only slightly slower than LCO, showing that the forward propagation through our neural networks in the maximizer is very fast. Note further that the solver in *scipy.optimize* package is much slower (i.e., $\sim 70x$) than our neural network. Thus, it is beyond our computational capability to use the solver along with an exhaustive search method – for each testing problem instance, approximately 2^{20} runs of the solver are needed to find the optimal robust decision in our experiment.

VI. CONCLUSION

In this work, we have studied robust combinatorial optimization and proposed LRCO, a learning-based optimizer that quickly outputs a robust solution in the presence of uncertain context. LRCO leverages a pair of learning-based optimizers — one for the minimizer and the other for the maximizer. Finally, to evaluate the performance of LRCO, we have performed simulations for the task offloading problem in vehicular edge computing. Our results highlighted that LRCO can greatly improve robustness, while having a very low runtime complexity.

ACKNOWLEDGEMENT

Z. Shaolei, J. Yang, and S. Ren are supported in part by the NSF under grants CNS-1551661 and CNS-1910208. C. Shen is supported in part by the NSF under grant SWIFT-2029978.

REFERENCES

- [1] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [2] O. Wichrowska, N. Maheswaranathan, M. W. Hoffman, S. G. Colmenarejo, M. Denil, N. Freitas, and J. Sohl-Dickstein, “Learned optimizers that scale and generalize,” in *ICML*, 2017.
- [3] H. Sun, X. Chen, Q. Shi, M. Hong, X. Fu, and N. D. Sidiropoulos, “Learning to optimize: Training deep neural networks for interference management,” *IEEE Transactions on Signal Processing*, vol. 66, no. 20, pp. 5438–5453, 2018.
- [4] F. Liang, C. Shen, W. Yu, and F. Wu, “Towards optimal power control via ensembling deep neural networks,” *IEEE Transactions on Communications*, vol. 68, no. 3, pp. 1760–1776, 2020.
- [5] Y. Xiong and C.-J. Hsieh, “Improved adversarial training via learned optimizer,” in *European Conference on Computer Vision*, 2020.
- [6] T. Chen, X. Chen, W. Chen, H. Heaton, J. Liu, Z. Wang, and W. Yin, “Learning to optimize: A primer and a benchmark,” 2021.
- [7] K. Li and J. Malik, “Learning to optimize,” in *ICLR*, 2017.
- [8] M. Andrychowicz, M. Denil, S. Gomez, M. W. Hoffman, D. Pfau, T. Schaul, B. Shillingford, and N. De Freitas, “Learning to learn by gradient descent by gradient descent,” in *NIPS*, 2016.
- [9] Y. Bengio, A. Lodi, and A. Prouvost, “Machine learning for combinatorial optimization: A methodological tour d’horizon,” *European Journal of Operational Research*, vol. 290, no. 2, pp. 405–421, 2021.
- [10] C. H. Papadimitriou and K. Steiglitz, *Combinatorial optimization: Algorithms and Complexity*. Courier Corporation, 1998.
- [11] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, “Edge computing: Vision and challenges,” *IEEE Internet of Things Journal*, vol. 3, pp. 637–646, Oct 2016.
- [12] N. Vesselinova, R. Steinert, D. F. Perez-Ramirez, and M. Boman, “Learning combinatorial optimization on graphs: A survey with applications to networking,” *IEEE Access*, vol. 8, pp. 120388–120416, June 2020.
- [13] D. Bhandari, C. Murthy, and S. K. Pal, “Genetic algorithm with elitist model and its convergence,” *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 10, no. 06, pp. 731–747, 1996.
- [14] V. Balakrishnan, S. Boyd, and S. Balemi, “Branch and bound algorithm for computing the minimum stability degree of parameter-dependent linear systems,” *Intl. J. of Robust and Nonlinear Control*, vol. 1, pp. 295–317, 1991.
- [15] I. Bello, H. Pham, Q. V. Le, M. Norouzi, and S. Bengio, “Neural combinatorial optimization with reinforcement learning,” in *ICLR Workshop*, 2017.
- [16] T. Barrett, W. Clements, J. Foerster, and A. Lvovsky, “Exploratory combinatorial optimization with reinforcement learning,” in *AAAI*, 2020.
- [17] E. Khalil, P. Le Bodic, L. Song, G. Nemhauser, and B. Dilkina, “Learning to branch in mixed integer programming,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2016.
- [18] H. Dai, E. B. Khalil, Y. Zhang, B. Dilkina, and L. Song, “Learning combinatorial optimization algorithms over graphs,” in *NIPS*, 2017.
- [19] H. Zhang, W. Li, S. Gao, X. Wang, and B. Ye, “Reles: A neural adaptive multipath scheduler based on deep reinforcement learning,” in *INFOCOM*, 2019.
- [20] F. Wang, C. Zhang, F. Wang, J. Liu, Y. Zhu, H. Pang, and L. Sun, “Deepcast: Towards personalized qoe for edge-assisted crowdcast with deep reinforcement learning,” *IEEE/ACM Transactions on Networking*, vol. 28, no. 3, pp. 1255–1268, 2020.
- [21] H. Mao, M. Schwarzkopf, S. B. Venkatakrishnan, Z. Meng, and M. Alizadeh, “Learning scheduling algorithms for data processing clusters,” in *SIGCOMM*, 2019.
- [22] Y. Chen, Y. Tan, and B. Zhang, “Exploiting vulnerabilities of load forecasting through adversarial attacks,” in *e-Energy*, 2019.
- [23] J. Shen, X. Chen, H. Heaton, T. Chen, J. Liu, W. Yin, and Z. Wang, “Learning a minimax optimizer: A pilot study,” in *ICLR*, 2021.
- [24] T. Chen, W. Zhang, Z. Jingyang, S. Chang, S. Liu, L. Amini, and Z. Wang, “Training stronger baselines for learning to optimize,” in *NeurIPS*, 2020.
- [25] Y. Chen, M. W. Hoffman, S. G. Colmenarejo, M. Denil, T. P. Lillicrap, M. Botvinick, and N. Freitas, “Learning to learn without gradient descent by gradient descent,” in *ICML*, 2017.
- [26] Y. Cao, T. Chen, Z. Wang, and Y. Shen, “Learning to optimize in swarms,” in *NeurIPS*, 2019.
- [27] H. Jiang, Z. Chen, Y. Shi, B. Dai, and T. Zhao, “Learning to defense by learning to attack,” in *AISTATS*, 2021.
- [28] B. Korte, J. Vygen, B. Korte, and J. Vygen, *Combinatorial optimization*. Springer, 2012.
- [29] O. Vinyals, M. Fortunato, and N. Jaitly, “Pointer networks,” in *NIPS*, 2015.
- [30] M. Deudon, P. Courtnut, A. Lacoste, Y. Adulyasak, and L.-M. Rousseau, “Learning heuristics for the TSP by policy gradient,” in *CPAIOR*, Springer, 2018.
- [31] I. Bogunovic, S. Mitrović, J. Scarlett, and V. Cevher, “Robust submodular maximization: A non-uniform partitioning approach,” 2017.
- [32] A. Nedić and A. Ozdaglar, “Subgradient methods for saddle-point problems,” *Journal of optimization theory and applications*, vol. 142, no. 1, pp. 205–228, 2009.
- [33] J. Hamm and Y.-K. Noh, “K-beam minimax: Efficient optimization for deep adversarial learning,” in *ICML*, 2018.
- [34] C. Daskalakis, A. Ilyas, V. Syrgkanis, and H. Zeng, “Training GANs with optimism,” *arXiv preprint arXiv:1711.00141*, 2017.
- [35] E. K. Ryu, K. Yuan, and W. Yin, “Ode analysis of stochastic gradient methods with optimism and anchoring for minimax problems and GANs,” *arXiv preprint arXiv:1905.10899*, 2019.
- [36] Y. Wang, G. Zhang, and J. Ba, “On solving minimax optimization locally: A follow-the-ridge approach,” in *ICLR*, 2020.
- [37] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” in *NIPS*, 2014.
- [38] J. B. Orlin, A. S. Schulz, and R. Udwani, “Robust monotone submodular function maximization,” *Mathematical Programming*, vol. 172, no. 1-2, pp. 505–537, 2018.
- [39] L. Huang, A. D. Joseph, B. Nelson, B. I. Rubinstein, and J. D. Tygar, “Adversarial machine learning,” in *AISeC*, 2011.
- [40] K. Hornik, M. Stinchcombe, and H. White, “Multilayer feedforward networks are universal approximators,” *Neural Netw.*, vol. 2, p. 359–366, July 1989.
- [41] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT Press, 2018.
- [42] G. Yan, D. Wen, S. Olariu, and M. C. Weigle, “Security challenges in vehicular cloud computing,” *IEEE Transactions on Intelligent Transportation Systems*, 2012.
- [43] L. Chen and J. Xu, “Task replication for vehicular cloud: Contextual combinatorial bandit with delayed feedback,” in *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*, pp. 748–756, IEEE, 2019.
- [44] “Global positioning system (gps) standard positioning service (sps) performance standard.” <https://www.gps.gov/technical/ps/2020-SPS-performance-standard.pdf>, 2020.
- [45] P. Purdue, “The m/m/1 queue in a markovian environment,” *Operations Research*, 1974.
- [46] S. Shekhar, A. D. Chhokra, A. Bhattacharjee, G. Aupy, and A. Gokhale, “Indices: exploiting edge resources for performance-aware cloud-hosted services,” in *2017 IEEE 1st International Conference on Fog and Edge Computing (ICFEC)*, 2017.
- [47] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.

APPENDIX CONTEXT PARAMETER PREDICTION

We consider two context parameter prediction models, which predict the success rate x based on the feature: $x = M_p(d, \text{cpu}, L)$. The first model is implemented with linear regression: $x = w_d d + w_{\text{cpu}} \text{cpu} + w_L L + \text{bias}$, which is easy to interpret. Additionally, we provide a neural network-based residual model, including the pre-trained linear model and a

parallel residual block. The residual block is implemented with a neural network to learn the residual error.

In the training phase, both linear and neural network models are trained with the same training dataset. Specifically, the neural network residual block includes two hidden layers, each with 20 neurons, and ‘*ReLU*’ activation. The residual block is trained by minimizing the ‘*mean squared error*’ loss function using the Adam optimizer (learning rate $1e - 4$).

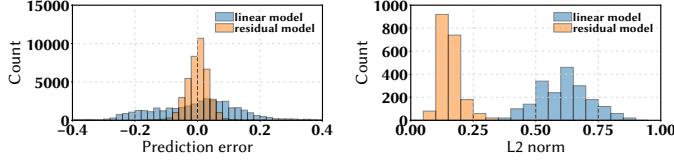


Fig. 8. Context parameter (success rate) estimation error and its L_2 norm.

The predicted success rate by the linear model is denoted as x^l , while x^r represents the result of the residual model. Then, two errors are investigated: the element-wise prediction error and the L_2 norm of 20 elements in the context parameter. We show the results in Fig. 8. The results validate that the residual model indeed reduces the estimation error. Thus, we will consider the robust optimization problem with both a large error budget and a small error budget. Here, we consider the 99-percentile L_2 norm as the error budget for robustness, which is 0.71 for the empirical model and 0.27 for the residual model, respectively.