ONLINE FACILITY LOCATION WITH PREDICTIONS

Anonymous authors
Paper under double-blind review

ABSTRACT

We provide nearly optimal algorithms for online facility location (OFL) with predictions. In OFL, n demand points arrive in order and the algorithm must irrevocably assign each demand point to an open facility upon its arrival. The objective is to minimize the total connection costs from demand points to assigned facilities plus the facility opening cost. We further assume the algorithm is additionally given for each demand point x_i a natural prediction $f_{x_i}^{\text{pred}}$ which is supposed to be the facility $f_{x_i}^{\text{opt}}$ that serves x_i in the offline optimal solution.

Our main result is an $O(\min\{\log \frac{n\eta_\infty}{\text{OPT}}, \log n\})$-competitive algorithm where η_∞ is the maximum prediction error (i.e., the distance between $f_{x_i}^{\text{pred}}$ and $f_{x_i}^{\text{opt}}$). Our algorithm overcomes the fundamental $\Omega(\frac{\log n}{\log\log n})$ lower bound of OFL (without predictions) when η_∞ is small, and it still maintains $O(\log n)$ ratio even when η_∞ is unbounded. Furthermore, our theoretical analysis is supported by empirical evaluations for the tradeoffs between η_∞ and the competitive ratio on various real datasets of different types.

1 INTRODUCTION

We study the online facility location (OFL) problem with predictions. In OFL, given a metric space $(\mathcal{X}, d)$, a ground set $\mathcal{D} \subseteq \mathcal{X}$ of demand points, a ground set $\mathcal{F} \subseteq \mathcal{X}$ of facilities, and an *opening cost* function $w : \mathcal{F} \to \mathbb{R}_+$, the input is a sequence of demand points $X := (x_1, \ldots, x_n) \subseteq \mathcal{D}$, and the online algorithm must *irrevocably* assign x_i to some *open* facility, denoted as f_i, upon its arrival, and an open facility cannot be closed. The goal is to minimize the following objective

$$\sum_{f \in F} w(f) + \sum_{x_i \in X} d(x_i, f_i),$$

where F is the set of open facilities.

Facility location is a fundamental problem in both computer science and operations research, and it has been applied to various domains such as supply chain management (Melo et al., 2009), distribution system design (Klose & Drexl, 2005), healthcare (Ahmadi-Javid et al., 2017), and more applications can be found in surveys (Drezner & Hamacher, 2002; Laporte et al., 2019). Its online variant, OFL, was also extensively studied and well understood. The state of the art is an $O(\frac{\log n}{\log\log n})$-competitive deterministic algorithm proposed by Fotakis (2008), and the same work shows this ratio is tight.

We explore whether or not the presence of certain natural predictions could help to bypass the $O(\frac{\log n}{\log\log n})$ barrier. Specifically, we consider the setting where each demand point x_i receives a predicted facility $f_{x_i}^{\text{pred}}$ that is supposed to be x_i's assigned facility in the (offline) optimal solution. This prediction is very natural, and it could often be easily obtained in practice. For instance, if the dataset is generated from a latent distribution, then predictions may be obtained by analyzing past data. Moreover, predictions could also be obtained from external sources, such as experts' advices, and additional features that define correlations among demand points and/or facilities. Previously, predictions of a similar flavor were also considered for other online problems, such as online caching (Rohatgi, 2020; Lykouris & Vassilvitskii, 2021), ski rental (Purohit et al., 2018; Gollapudi & Panigrahi, 2019) and online revenue maximization (Medina & Vassilvitskii, 2017).

1.1 Our Results

Our main result is a near-optimal online algorithm that offers a smooth tradeoff between the prediction error and the competitive ratio. In particular, our algorithm is $O(1)$-competitive when the predictions are perfectly accurate, i.e., $f_{x_i}^{\text{pred}}$ is the nearest neighbor of x_i in the optimal solution for every $1 \le i \le n$. On the other hand, even when the predictions are completely wrong, our algorithm still remains $O(\log n)$-competitive. As in the literature of online algorithms with predictions, our algorithm does not rely on the knowledge of the prediction error η_∞.

Theorem 1.1. *There exists an $O(\min\{\log \frac{n\eta_\infty}{\text{OPT}}, \log n\})$-competitive algorithm for the OFL with predictions, where $\eta_\infty := \max_{1\le i\le n} d(f_{x_i}^{\text{pred}}, f_{x_i}^{\text{opt}})$ measures the maximum prediction error, and $f_{x_i}^{\text{opt}}$ is the nearest neighbor of x_i from the offline optimal solution* OPT.

Indeed, we can also interpret our result under the robustness-consistency framework which is widely considered in the literature of online algorithms with predictions (cf. Lykouris & Vassilvitskii (2021)), where the robustness and consistency stand for the competitive ratios of the algorithm in the cases of arbitrarily bad predictions and perfect predictions, respectively. Under this language, our algorithm is $O(1)$-consistent and $O(\log n)$-robust. We also remark that our robustness bound nearly matches the lower bound for the OFL without predictions.

One might wonder whether or not it makes sense to consider a related error measure, $\eta_1 := \sum_{i=1}^n d(f_{x_i}^{\text{pred}}, f_{x_i}^{\text{opt}})$, which is the *total* error of predictions. Our lower bound result (Theorem 1.2) shows that a small η_1 might not be helpful, and the dependence of η_∞ in our upper bound is fundamental. The lower bound asserts that our algorithm is nearly tight in the dependence of η_∞.

Theorem 1.2. *For every η_∞, there exists a class of inputs, such that every (randomized) online algorithm is $\Omega(\log \frac{n\eta_\infty}{\text{OPT}})$-competitive, even when $\forall f \in \mathcal{F}, w(f) = 1$, and $\eta_1 = O(1)$.*

Empirical evaluation. We simulate our algorithm on both Euclidean and graph (shortest-path) datasets, and we measure the tradeoff between η_∞ and the empirical competitive ratio, by generating random predictions whose η_∞ is controlled to be around some target value. We compare with two baselines, the $O(\log n)$-competitive algorithm by Meyerson (2001) which do not use the prediction at all, and a naive algorithm that always trusts the prediction. Our algorithm significantly outperforms Meyerson's algorithm when $\eta_\infty \to 0$, and is still comparable to Meyerson's algorithm when η_∞ is large where the follow-prediction baseline suffers a huge error.

We observe that our lead is even more significant on datasets with non-uniform opening cost, and this suggests that the prediction could be very useful in the non-uniform setting. This phenomenon seems to be counter-intuitive since the error guarantee η_∞ only concerns the connection costs without taking the opening cost into consideration (i.e., it could be that a small η_∞ is achieved by predictions of huge opening cost), which seems to mean the prediction may be less useful. Therefore, this actually demonstrates the superior capability of our algorithm in using the limited information from the predictions even in the non-uniform opening cost setting.

Comparison to an independent work. A recent independent work by Fotakis et al. (2021a) considers OFL with predictions in a similar setting. We highlight the key differences as follows.

(i) We allow arbitrary opening cost $w(\cdot)$, while Fotakis et al. (2021a) only solves the special case of *uniform* facility location where all opening costs are equal. This is a significant difference, since as we mentioned (and will discuss in Section 1.2)), the non-uniform opening cost setting introduces outstanding technical challenges that require novel algorithmic ideas.

(ii) For the uniform facility location setting, our competitive ratio is $O(\log \frac{n\eta_\infty}{\text{OPT}}) = O(\log \frac{n\eta_\infty}{w})$, while the ratio in Fotakis et al. (2021a) is $\frac{\log(\text{err}_0)}{\log(\text{err}_1^{-1}\log(\text{err}_0))}$, where $\text{err}_0 = n\eta_\infty/w$, $\text{err}_1 = \eta_1/w$ and w is the uniform opening cost. Our ratio is comparable to theirs when η_1 is relatively small. However, theirs seems to be better when η_1 is large, and in particular, it was claimed in Fotakis et al. (2021a) that the ratio becomes constant when $\text{err}_1 \approx \text{err}_0$. Unfortunately, we think this claim contradicts with our lower bound (Theorem 1.2) which essentially asserts that $\eta_1 = O(\eta_\infty)$ is not helpful for the ratio. Moreover, when η_1 is large, we find their ratio is actually not well-defined since the denominator $\log(\text{err}_1^{-1}\log(\text{err}_0))$ is negative. Therefore, we believe there might be some unstated technical assumptions, or there are technical issues remaining to be fixed, when η_1 is large.

1.2 Technical Overview

Since we aim for a worst-case (i.e., when the predictions are inaccurate) $O(\log n)$ ratio, our algorithm is based on an $O(\log n)$-competitive algorithm (that does not use predictions) by Meyerson (2001), and we design a new procedure to make use of predictions. In particular, upon the arrival of each demand point, our algorithm first runs the steps of Meyerson's algorithm , which we call the Meyerson step, and then runs a new procedure (proposed in this paper) to open additional facilities that are "near" the prediction, which we call the Prediction step.

We make use of the following crucial property of Meyerson's algorithm. Suppose before the first demand point arrives, the algorithm is already provided with an initial facility set $\bar{F}$, such that for every facility f^* opened in OPT, $\bar{F}$ contains a facility f' that is of distance η_∞ to f^*, then Meyerson's algorithm is $O(\frac{n\eta_\infty}{\mathrm{OPT}})$-competitive. To this end, our Prediction step aims to open additional facilities so that for each facility in OPT, the algorithm would soon open a close enough facility.

Uniform opening cost. There is a simple strategy that achieves this goal for the case of *uniform opening cost*. Whenever the Meyerson step decides to open a facility, we further open a facility at the predicted location. By doing so, we would pay twice as the Meyerson algorithm does in order to open the extra facility. As a reward, when the prediction error η_∞ is small, the extra facility would be good and close enough to the optimal facility.

Non-uniform opening cost. Unfortunately, this strategy does not extend to the non-uniform case. Specifically, opening a facility exactly at the predicted location could be prohibitively expensive as it may have huge opening cost. Instead, one needs to open a facility that is "close" to the prediction, but attention must be paid to the tradeoff between the opening cost and the proximity to the prediction.

The design of the Prediction step for the non-uniform case is the most technical and challenging of our paper. We start with opening an initial facility that is far from the prediction, and then we open a series of facilities within balls (centered at the prediction) of geometrically decreasing radius, while we also allow the opening cost of the newly open facility to be doubled each time. We stop opening facilities if the total opening cost exceeds the cost incurred by the preceding Meyerson step, in order to have the opening cost bounded.

We show that our procedure always opens a "correct" facility $\hat{f}$ that is $\Theta(\eta_\infty)$ apart to the corresponding facility in OPT, and that the total cost until $\hat{f}$ is opened is merely $O(\mathrm{OPT})$. This is guaranteed by the gradually decreasing ball radius when we build additional facilities (so that the "correct" facility will not be skipped), and that the total opening cost could be bounded by that of the last opened facility (because of the doubling opening cost). Once we open this $\hat{f}$, subsequent runs of Meyerson steps would be $O(\log \frac{n\eta_\infty}{\mathrm{OPT}})$-competitive.

1.3 Related Work

Online facility location problem is introduced by Meyerson (2001). Meyerson introduced a randomized algorithm that achieves competitive ratios of $O(1)$ and $O(\log n)$ in the setting of random arrival order and the setting of adversarial arrival order respectively. Later, Fotakis (2008) gave a lower bound of $\Omega(\frac{\log n}{\log\log n})$ and proposed a determined algorithm matching the lower bound. Fotakis also improved the analysis of Meyerson's algorithm, by showing an $\Omega(\frac{\log n}{\log\log n})$ competitive ratio in the setting of adversarial order. Other variants of the online facility location problem are studied as well. Divéki & Imreh (2011) and Bamas et al. (2020b) considered the dynamic variant, in which an open facility can be reassigned. Online facility location with evolving metrics was studied by Eisenstat et al. (2014) and Fotakis et al. (2021b).

There is a recent trend of studying online algorithms with predictions and many classical online problems have been revisited in this framework. Lykouris & Vassilvitskii (2021) considered online caching problem with predictions and gave a formal definition of consistency and robustness. For the ski-rental problem, Purohit et al. (2018) gave an algorithm with a hyper parameter to maintain the balance of consistency and robustness. Gollapudi & Panigrahi (2019) designed an algorithm using multiple predictors to achieve low prediction error. Recently, Wei & Zhang (2020) showed a tight robustness-consistency tradeoff for the ski-rental problem. For the caching problem, Lykouris

& Vassilvitskii (2021) adapted Marker algorithm to use predictions. Following this work, Rohatgi (2020) provided an improved algorithm that performs better when the predictor is misleading. Bamas et al. (2020b) extended the online primal-dual framework to incorporate predictions and applied their framework to solve the online covering problem. Medina & Vassilvitskii (2017) considered the online revenue maximization problem and Bamas et al. (2020a) studied the online speed scaling problem. Antoniadis et al. (2020), Jiang et al. (2021) studied the secretary and online matching problems with predictions in the robustness-consistency framework and the constrained adversary framework respectively.

2 Preliminaries

Recall that we assume a underlying metric space $(\mathcal{X}, d)$. For $S \subseteq \mathcal{X}, x \in \mathcal{X}$, define $d(x, S) := \min_{y \in S} d(x, y)$. For an integer t, let $[t] := \{1, 2, \ldots, t\}$.

We normalize the opening cost function w so that the minimum opening cost equals 1, and we round down the opening cost of each facility to the closest power of 2. This only increases the ratio by a factor of 2 which we can afford (since we make no attempt to optimize the constant hidden in big O). Hence, we assume the domain of w is $\{2^{i-1} \mid i \in [L]\}$ for some L. Further, we use G_k to denote the set of facilities whose opening cost is at most 2^{k-1}, i.e. $G_k := \{f \in \mathcal{F} \mid w(f) \leq 2^{k-1}\}, \forall 0 \leq k \leq L$. Observe that $G_0 = \emptyset$.

3 Our Algorithm: Prediction-augmented Meyerson

Our algorithm consists of two steps, the Meyerson step and the Prediction step. Upon the arrival of each demand point x, we first run the Meyerson algorithm Meyerson (2001), and let $\mathrm{cost}_{\mathrm{M}}(x)$ be the cost from the Meyerson step. Next, we utilize the prediction information by opening more facilities in the Prediction step, with a total budget of $\mathrm{cost}_{\mathrm{M}}(x)$.

Algorithm 1 Prediction-augmented Meyerson

initialize $F \leftarrow \emptyset$, $F_{\mathrm{P}} \leftarrow \emptyset$ ▷ F, F_{P} are the set of all open facilities and those opened by PRED
for arriving demand point $x \in X$ and its associated f_x^{pred} **do**
  $\mathrm{cost}_{\mathrm{M}}(x) \leftarrow$ MEY (x)
  PRED $\left(f_x^{\mathrm{pred}}, \mathrm{cost}_{\mathrm{M}}(x)\right)$
procedure MEY(x)
  for $k \in [L]$ **do**
    $f_k \leftarrow \arg\min_{f \in F \cup G_k} d(x, f)$
    $\delta_k \leftarrow d(x, f_k)$
    $p_k \leftarrow (\delta_{k-1} - \delta_k)/2^k$, where $\delta_0 = d(x, F)$
  let $s_k \leftarrow \sum_{i=k}^{L} p_i$ for $k \in [L]$
  sample $r \sim \mathrm{Uni}[0, 1]$, and let $i \in [L]$ be the index such that $r \in [s_{i+1}, s_i)$
  if such i exists **then**
    $F \leftarrow F \cup f_i$ ▷ open facility at f_i
  connect x to the nearest neighbor in F
  return $d(x, F) + w(f_i)$ ▷ if i does not exist, simply return $d(x, F)$
procedure PRED(f_x^{pred}, q)
  repeat
    $r \leftarrow \frac{1}{2} d(f_x^{\mathrm{pred}}, F_{\mathrm{P}})$
    $f_{\mathrm{open}} \leftarrow \arg\min_{f: d(f_x^{\mathrm{pred}}, f) \leq r} w(f)$ ▷ break ties by picking the closest facility to f_x^{pred}
    if $q \geq w(f_{\mathrm{open}})$ **then**
      $F \leftarrow F \cup f_{\mathrm{open}}$, $F_{\mathrm{P}} \leftarrow F_{\mathrm{P}} \cup f_{\mathrm{open}}$ ▷ open facility at f_{open}
      $q \leftarrow q - w(f_{\mathrm{open}})$
  until $q < w(f_{\mathrm{open}})$
  open facility at f_{open} with probability $\frac{q}{w(f_{\mathrm{open}})}$, and update F, F_{P}

Theorem 3.1. *Prediction-augmented Meyerson is $O(\min\{\log \frac{n\eta_\infty}{\mathrm{OPT}}, \log n\})$-competitive.*

Calibrating predictions. Recall that $\eta_\infty = \max_{i\in[n]} d(f_{x_i}^{\text{pred}}, f_{x_i}^{\text{opt}})$ can be unbounded. We show how to "calibrate" the bad predictions so that $\eta_\infty = O(\text{OPT})$. Specifically, when a demand point x arrives, we compute $f'_x := \arg\min_{f\in\mathcal{F}}\{d(x,f) + w(f)\}$, and we calibrate f_x^{pred} by letting $f_x^{\text{pred}} = f'_x$ if $d(x, f_x^{\text{pred}}) \geq 2d(x, f'_x) + w(f'_x)$.

We show the calibrated predictions satisfy $\eta_\infty = O(\text{OPT})$. Indeed, for every demand point x, $d(f'_x, f_x^{\text{opt}}) \leq d(x, f'_x) + d(x, f_x^{\text{opt}}) \leq 2d(x, f'_x) + w(f'_x) \leq O(\text{OPT})$, where the second last inequality follows from the optimality (i.e., the connection cost $d(x, f_x^{\text{opt}})$ has to be smaller than the cost of first opening f'_x then connecting x to f'_x). Note that, if $\eta_\infty = O(\text{OPT})$ then $\frac{n\eta_\infty}{\text{OPT}} = O(\log n)$, hence it suffices to prove a single bound $O(\log \frac{n\eta_\infty}{\text{OPT}})$ for Theorem 3.1.

Algorithm analysis. Let F_{opt} be the set of open facilities in the optimal solution. We examine each $f^* \in F_{\text{opt}}$ and its corresponding demand points separately, i.e. those demand points connecting to f^* in the optimal solution. We denote this set of demand points by $X(f^*)$.

Definition 3.2 (Open facilities)**.** For every demand point x, let $F(x)$ be the set of open facilities right after the arrival of request x, $\bar{F}(x)$ be the set of open facilities right before the arrival of x, and $\hat{F}(x) = F(x) \setminus \bar{F}(x)$ be the set of the newly-open facilities on the arrival of x. Moreover, let $F_{\text{M}}(x), F_{\text{P}}(x)$ be a partition of $F(x)$, corresponding to the facilities opened by MEY and PRED, respectively. Let $\bar{F}_{\text{M}}(x), \hat{F}_{\text{M}}(x), \bar{F}_{\text{P}}(x), \hat{F}_{\text{P}}(x)$ be defined similarly.

Definition 3.3 (Costs)**.** Let $\text{cost}_{\text{M}}(x) = w(\hat{F}_{\text{M}}(x)) + d\left(x, \left(F_{\text{M}}(x) \cup \bar{F}_{\text{P}}(x)\right)\right)$ be the the total cost from MEY step. Let $\text{cost}_{\text{P}}(x) = w(\hat{F}_{\text{P}}(x))$ be the cost from PRED step. Let $\text{cost}(x) = \text{cost}_{\text{M}}(x) + \text{cost}_{\text{P}}(x)$ be the total cost of x.

Recall that after the MEY step, we assign a total budget of $\text{cost}_{\text{M}}(x)$ to the PRED step. That is, the expected cost from the PRED step is upper bounded by the cost from the MEY step. We formalize this intuition in the following lemma.

Lemma 3.4. *For each demand point $x \in X$, $\mathbb{E}[\text{cost}(x)] = 2\,\mathbb{E}[\text{cost}_{\text{M}}(x)]$.*

Proof. Consider the expected cost from the PRED step. Given arbitrary F_p, q, there is only one random event from the algorithm and it is straightforward to see that the expected cost equals q. Notice that our algorithm assigns a total budget of $q = \text{cost}_{\text{M}}(x)$. Thus, $\mathbb{E}[\text{cost}(x)] = \mathbb{E}[\text{cost}_{\text{M}}(x) + \text{cost}_{\text{P}}(x)] = 2\,\mathbb{E}[\text{cost}_{\text{M}}(x)]$. □

Therefore, we are left to analyze the total expected cost from the MEY step. The following lemma is the most crucial to our analysis. Before continuing to the proof of the lemma, we explain how it concludes the proof of our main theorem.

Lemma 3.5. *For every facility $f^* \in F_{\text{opt}}$, we have*

$$\sum_{x\in X(f^*)} \mathbb{E}[\text{cost}_{\text{M}}(x)] \leq O\left(\log \frac{n\eta_\infty}{\text{OPT}}\right) \cdot \left(w(f^*) + \sum_{x\in X(f^*)} d(x, f^*)\right) + O\left(\frac{|X(f^*)|}{n} \cdot \text{OPT}\right).$$

Proof of Theorem 1.1. Summing up the equation of Lemma 3.5 for every $f^* \in F_{\text{opt}}$, we have

$$\begin{aligned}
\mathbb{E}[\text{ALG}] &= \sum_{x\in X} \mathbb{E}[\text{cost}(x)] = 2\sum_{x\in X} \mathbb{E}[\text{cost}_{\text{M}}(x)] = 2\sum_{f^*\in F_{\text{opt}}} \sum_{x\in X(f^*)} \mathbb{E}[\text{cost}_{\text{M}}(x)] \\
&\leq \sum_{f^*\in F_{\text{opt}}} \left(O\left(\log \frac{n\eta_\infty}{\text{OPT}}\right) \cdot \left(w(f^*) + \sum_{x\in X(f^*)} d(x, f^*)\right) + O\left(\frac{|X(f^*)|}{n} \cdot \text{OPT}\right)\right) \\
&= O\left(\log \frac{n\eta_\infty}{\text{OPT}}\right) \cdot \text{OPT},
\end{aligned}$$

where the second equality is by Lemma 3.4 and the inequality in by Lemma 3.5. □

3.1 PROOF OF LEMMA 3.5

Fix an arbitrary $f^* \in F_{\text{opt}}$. Let ℓ^* be the integer that $w(f^*) = 2^{\ell^*-1}$, or equivalently $f^* \in G_{\ell^*}$. Let $X(f^*) = \{x_1, x_2, \ldots, x_m\}$ be listed according to their arrival order. We remark that there can be other demand points arriving between x_i and x_{i+1}, but they must be connected to other facilities in the optimal solution. Let ℓ_i be the index ℓ that $d(f^*, \bar{F}(x_i)) \in [2^{\ell-1}, 2^{\ell})$. ℓ_i can be negative and if $d(f^*, \bar{F}(x_i)) = 0$, let $\ell_i = -\infty$. Observe that $d(f^*, \bar{F}(x_i))$ and ℓ_i are non-increasing in i. Let τ be the largest integer i with $d(f^*, \bar{F}(x_i)) > \min\{7\eta_\infty, 4w(f^*)\}$. Note that τ is a random variable. We refer to the demand sequence before x_τ as the long-distance stage and the sequence after x_τ as the short-distance stage.

In this section, we focus on the case when $7\eta_\infty \leq 4w(f^*)$. The proof for the other case is very similar which can be found in Section B.

Long-distance stage. We start with bounding the total cost in the long-distance stage. Intuitively, our procedure PRED would quickly build a facility that is close to f^* within a distance of $O(\eta_\infty)$, since it is guaranteed that for each demand point x_i, the distance between f^* and its corresponding prediction $f_{x_i}^{\text{pred}}$ is at most η_∞. Formally, we prove the following statements.

Lemma 3.6. $\mathbb{E}\left[\sum_{i<\tau} \text{cost}_{\text{M}}(x_i)\right] \leq 2w(f^*)$.

Proof. Let $G = \{g_1, g_2, \ldots, g_t\}$ be the facilities opened by the PRED steps on the arrivals of all $\{x_i\}_{i<\tau}$, and are listed according to their opening order. Observe that the budget that is assigned to the PRED step is $\text{cost}_{\text{M}}(x_i)$, and the expected opening cost of each PRED step equals its budget. By the definition of G, we have

$$\mathbb{E}\left[\sum_{i<\tau} \text{cost}_{\text{M}}(x_i)\right] = \mathbb{E}\left[\sum_{k\in[t]} w(g_k)\right].$$

Next, we prove that the opening costs of g_k's are always strictly increasing. For each $k \in [t-1]$, suppose g_k is opened on the arrival of demand x and g_{k+1} is opened on the arrival of demand y. Let F denote the set of all open facilities right before g_{k+1} is opened. Let $f := \arg\min_{f: d(f_y^{\text{pred}}, f) \leq \frac{1}{2} d(f_y^{\text{pred}}, g_k)} w(f)$, we must have $w(g_{k+1}) \geq w(f)$, since $g_{k+1} = \arg\min_{f: d(f_y^{\text{pred}}, f) \leq \frac{1}{2} d(f_y^{\text{pred}}, F)} w(f)$ and that $d(f_y^{\text{pred}}, F) \leq d(f_y^{\text{pred}}, g_k)$. Moreover,

$$\begin{aligned} d(f, f_x^{\text{pred}}) &\leq d(f, f_y^{\text{pred}}) + d(f_x^{\text{pred}}, f_y^{\text{pred}}) \leq \frac{1}{2} d(f_y^{\text{pred}}, g_k) + 2\eta_\infty \\ &\leq \frac{1}{2}\left(d(f_x^{\text{pred}}, g_k) + d(f_x^{\text{pred}}, f_y^{\text{pred}})\right) + 2\eta_\infty \leq \frac{1}{2} d(f_x^{\text{pred}}, g_k) + 3\eta_\infty < d(f_x^{\text{pred}}, g_k). \end{aligned}$$

Here, the second and fourth inequalities use the fact that $d(f_x^{\text{pred}}, f_y^{\text{pred}}) \leq d(f_x^{\text{pred}}, f^*) + d(f^*, f_y^{\text{pred}}) \leq 2\eta_\infty$ since both $x, y \in X(f^*)$. The last inequality follows from the fact that $d(f_x^{\text{pred}}, g_k) \geq d(g_k, f^*) - d(f_x^{\text{pred}}, f^*) > 7\eta_\infty - \eta_\infty = 6\eta_\infty$ by the definition of τ.

Consider the moment when we open g_k, the fact that we does not open facility f implies that $w(g_k) < w(f)$. Therefore, $w(g_k) < w(f) \leq w(g_{k+1})$. Recall that the opening cost of each facility is a power of 2. We have that $\sum_{k\in[t]} w(g_k) \leq 2w(g_t)$.

Finally, we prove that the opening cost $w(g_t)$ is at most $w(f^*)$. Consider the moment when g_t is open, let x be the corresponding demand point. Notice that

$$d(f^*, f_x^{\text{pred}}) \leq \eta_\infty < \frac{1}{2} d(f^*, \bar{F}(x)) \leq \frac{1}{2} d(f^*, \bar{F}_{\text{P}}(x)),$$

by the definition of τ. However, we open g_t instead of f^* on Line 20 of our algorithm. It must be the case that $w(f^*) \geq w(g_t)$.

To conclude the proof, we have $\sum_{k\in[t]} w(g_k) \leq 2w(g_t) \leq 2w(f^*)$. □

Lemma 3.7. $\mathbb{E}[\text{cost}_{\text{M}}(x_\tau)] \leq 2w(f^*) + 2\sum_{i=1}^{m} d(x_i, f^*)$.

Proof. We bound the opening cost and connection cost separately.

Opening cost. By the algorithm, we know for every $1 \le i \le m$,

$$\mathbb{E}\left[w(\hat{F}_{\mathrm{M}}(x_i))\mathbf{1}(w(\hat{F}_{\mathrm{M}}(x_i)) > w(f^*))\right] \le \sum_{i>l^*} \frac{\delta_{i-1}-\delta_i}{2^i} \cdot 2^i = \delta_{l^*} \le d(x_i, f^*).$$

Hence, using the fact that at most one facility is opened by Meyerson step for each demand,

$$\begin{aligned}\mathbb{E}[w(\hat{F}_{\mathrm{M}}(x_\tau))] &\le \mathbb{E}\left[\max_{f\in F_{\mathrm{M}}(x_m)} w(f)\right] \le w(f^*) + \mathbb{E}\left[\sum_{f\in F_{\mathrm{M}}(x_m), w(f)>w(f^*)} w(f)\right] \\ &\le w(f^*) + \sum_{i=1}^{m} d(x_i, f^*).\end{aligned}$$

This finishes the analysis for the opening cost.

Connection cost. We claim that $d(x_\tau, F(x_\tau)) \le w(f^*) + d(x_\tau, f^*)$ (with probability 1).

Let $\delta_0, \delta_1, \ldots, \delta_L$ be defined as in our algorithm. We assume $\delta_0 = d(x_\tau, \bar{F}(x_\tau)) > w(f^*) + d(x_\tau, f^*)$, as otherwise the claim holds immediately. Let $k := \max\{k : \delta_k > w(f^*) + d(x_\tau, f^*)\}$. Then

$$\begin{aligned}&\Pr[d(x_\tau, F(x_\tau)) \le w(f^*) + d(x_\tau, f^*)] \\ &= \min\left(\sum_{i=k+1}^{L} \frac{\delta_{i-1}-\delta_i}{2^i}, 1\right) \ge \min\left(\sum_{i=k+1}^{\ell^*} \frac{\delta_{i-1}-\delta_i}{2^{\ell^*}}, 1\right) = \min\left(\frac{\delta_k - \delta_{\ell^*}}{2^{\ell^*}}, 1\right) = 1.\end{aligned}$$

Therefore, $\mathbb{E}[\mathrm{cost}_{\mathrm{M}}(x_\tau)] = \mathbb{E}[w(\hat{F}_{\mathrm{M}}(x_\tau))] + \mathbb{E}[d(f^*, F(x_\tau))] \le 2w(f^*) + 2\sum_{i=1}^{m} d(x_i, f^*)$, which concludes the proof. □

Short-distance stage. Next, we bound the costs for the subsequence of demand points whose arrival is after the event that some facility near f^* is open. Formally, these demand points are x_i's with $i > \tau$. The analysis of this part is conceptually similar to a part of the analysis of Meyerson's algorithm. However, ours is more refined in that it utilizes the already opened facility that is near f^* to get an improved $O(\log \frac{n\eta_\infty}{\mathrm{OPT}})$ ratio instead of $O(\log n)$. Moreover, Meyerson did not provide the full detail of this analysis, and our complete self-contained proof fills in this gap.

By the definition of τ, we have $d(f^*, \bar{F}(x_i)) \le 7\eta_\infty$ for such $i > \tau$. For integer ℓ, let $I_\ell = \{i > \tau \mid \ell_i = \ell\}$. Then for t such that $2^{t-1} > \min\{7\eta_\infty, 4w(f^*)\}$, we have $I_t = \emptyset$. Let $\underline{\ell}$ be the integer that $\frac{\mathrm{OPT}}{n} \in [2^{\underline{\ell}-1}, 2^{\underline{\ell}})$, and let $\bar{\ell}$ be the integer that $\min\{7\eta_\infty, 4w(f^*)\} \in [2^{\bar{\ell}-1}, 2^{\bar{\ell}})$. We partition all demand points with $i > \tau$ into groups $I_{\le\underline{\ell}-1}, I_{\underline{\ell}}, \ldots, I_{\bar{\ell}}$ according to ℓ_i, where $I_{\le\underline{\ell}-1} = \bigcup_{\ell\le\underline{\ell}-1} I_\ell$.

Lemma 3.8. $\mathbb{E}\left[\sum_{i\in I_{\le\underline{\ell}-1}} \mathrm{cost}_{\mathrm{M}}(x_i)\right] \le \frac{2|X(f^*)|}{n} \cdot \mathrm{OPT}$.

Proof. For $i \in I_{\le\underline{\ell}-1}$, we have that its connection cost $d(x_i, F(x_i)) \le d(x_i, \bar{F}(x_i)) \le 2^{\underline{\ell}-1} \le \frac{\mathrm{OPT}}{n}$. Let $\delta_0, \delta_1, \ldots, \delta_L$ and $f_1, \ldots, f_L$ be defined as in our algorithm upon x_i's arrival. Then, its expected opening cost in the MEY step equals

$$\begin{aligned}\mathbb{E}\left[w(\hat{F}_{\mathrm{M}}(x_i))\right] &= \sum_{k\in[L]} \Pr\left[\hat{F}_{\mathrm{M}}(x_i) = \{f_k\}\right] \cdot w(f_k) \le \sum_{k\in[L]} \min(p_k, 1) \cdot 2^{k-1} \\ &\le \sum_{k\in[L]} \min\left(\frac{\delta_{k-1}-\delta_k}{2^k}, 1\right) \cdot 2^{k-1} \le \sum_{k\in[L]} \frac{\delta_{k-1}-\delta_k}{2} \le \frac{\delta_0}{2} \le \frac{\mathrm{OPT}}{2n}.\end{aligned}$$

To sum up,

$$\begin{aligned}\mathbb{E}\left[\sum_{i\in I_{\le\underline{\ell}-1}} \mathrm{cost}_{\mathrm{M}}(x_i)\right] &\le \mathbb{E}\left[\sum_{i\in I_{\le\underline{\ell}-1}} \left(d(x_i, F(x_i)) + w(\hat{F}_{\mathrm{M}}(x_i))\right)\right] \le \sum_{i\in I_{\le\underline{\ell}-1}} \frac{3}{2n}\mathrm{OPT} \\ &< \frac{2|X(f^*)|}{n} \cdot \mathrm{OPT}.\end{aligned}$$

□

Lemma 3.9. *For every* $\ell \in [\underline{\ell}, \overline{\ell}]$, $\mathbb{E}\left[\sum_{i \in I_\ell} \mathrm{cost}_{\mathrm{M}}(x_i)\right] \leq 18 \sum_{i \in I_\ell} d(x_i, f^*) + 32w(f^*)$.

Proof. The proof can be found in Section A. □

Proof of Lemma 3.5. We conclude Lemma 3.5 by combining Lemma 3.6, 3.7, 3.8, and 3.9.

$$\begin{aligned}
\sum_{i \in [m]} \mathbb{E}\left[\mathrm{cost}_{\mathrm{M}}(x_i)\right] &= \mathbb{E}\left[\sum_{i \in [\tau-1]} \mathrm{cost}_{\mathrm{M}}(x_i)\right] + \mathbb{E}[\mathrm{cost}_{\mathrm{M}}(x_\tau)] + \sum_{\ell=\underline{\ell}}^{\overline{\ell}} \mathbb{E}\left[\sum_{i \in I_\ell} \mathrm{cost}_{\mathrm{M}}(x_i)\right] \\
&\quad + \sum_{i \in I_{\leq \underline{\ell}-1}} \mathbb{E}\left[\mathrm{cost}_{\mathrm{M}}(x_i)\right] \\
&\leq 2w(f^*) + \left(2w(f^*) + 2\sum_{i=1}^{m} d(x_i, f^*)\right) + \sum_{\ell=\underline{\ell}}^{\overline{\ell}} \left(18 \sum_{i \in I_\ell} d(x_i, f^*) + 32w(f^*)\right) \\
&\quad + \frac{2|X(f^*)|}{n} \cdot \mathrm{OPT} \\
&\leq O(\overline{\ell} - \underline{\ell}) \cdot \left(\sum_{i \in [m]} d(x_i, f^*) + w(f^*)\right) + \frac{2|X(f^*)|}{n} \mathrm{OPT} \\
&\leq O\left(\log \frac{n\eta_\infty}{\mathrm{OPT}}\right) \cdot \left(\sum_{i \in [m]} d(x_i, f^*) + w(f^*)\right) + O\left(\frac{|X(f^*)|}{n} \mathrm{OPT}\right).
\end{aligned}$$

□

4 Experiments

We validate our online algorithms on various datasets of different types. In particular, we consider three Euclidean data sets, a) Twitter (Chan et al.), b) Adult (Dua & Graff, 2017), and c) Non-Uni (Cebecauer & Buzna, 2018) which are datasets consisting of numeric features in $\mathbb{R}^d$ and the distance is measured by ℓ_2, and one graph dataset, US-PG (Rossi & Ahmed, 2015) which represents US power grid in a graph, where the points are vertices in a graph and the distance is measured by the shortest path distance. The opening cost is non-uniform in Non-Uni dataset, while it is uniform in all the other three. A summary of the specification of datasets can be found in Table 1.

Table 1: Specifications of datasets

dataset	type	size	# of dimension/edges	non-uniform
Twitter	Euclidean	30k	# dimension = 2	no
Adult	Euclidean	32k	# dimension = 6	no
US-PG	Graph	4.9k	# edges = 6.6k	no
Non-Uni	Euclidean	4.8k	# dimension = 2	yes

Our experiment aims to measure the tradeoff between η_∞ and the empirical competitive ratio of our algorithm, and we compare against two "extreme" baselines, a) the vanilla Meyerson's algorithm *without using* predictions which we call Meyerson, and b) a naive algorithm that *always follows* the prediction which we call Follow-Predict. Intuitively, a simultaneously consistent and robust online algorithm should perform similarly to the Follow-Predict baseline when the prediction is nearly perfect, while comparable to Meyerson when the prediction is of low quality.

Since it is NP-hard to find the optimal solution to facility location problem, we run a simple 3-approximate MP algorithm (Mettu & Plaxton, 2000) to find a near optimal solution $F^\star$ for every dataset, and we use this solution as the offline optimal solution (which we use as the benchmark for the competitive ratio). Then, for a given η_∞ and a demand point x, we pick a random facility $f \in \mathcal{F}$ such that $\eta_\infty/2 \leq d(f, F_x^\star) \leq \eta_\infty$ as the prediction, where $F_x^\star$ is the point in $F^\star$ that is closest to

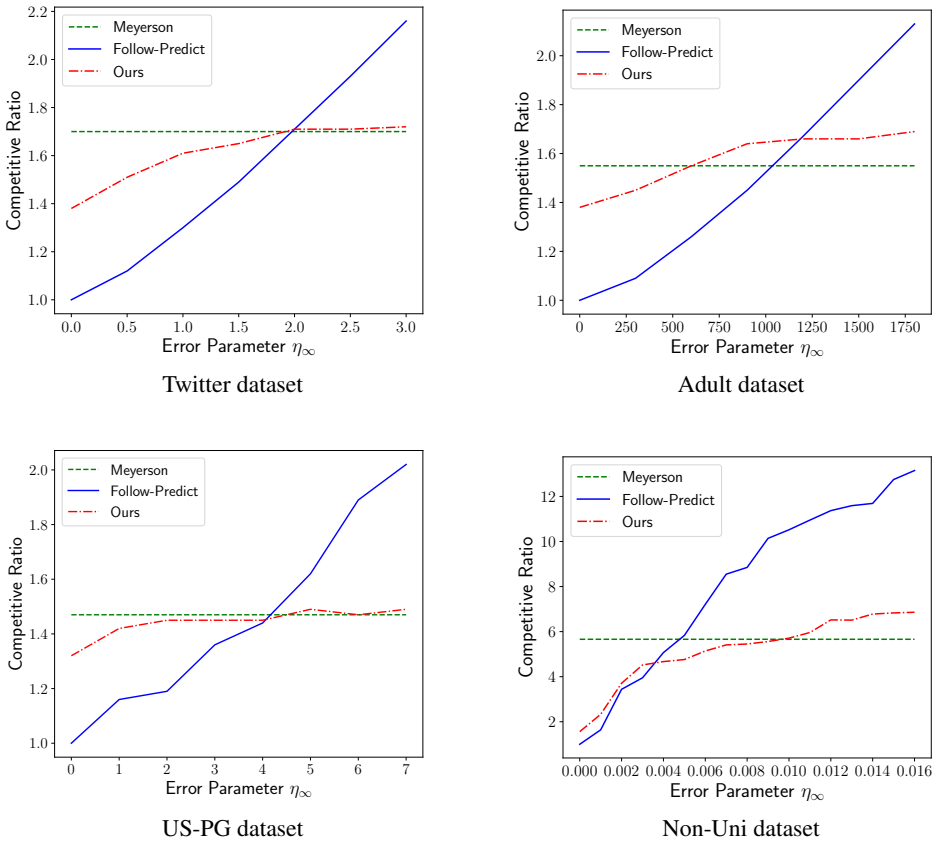

Figure 1: The tradeoff between prediction error η_∞ and empirical competitive ratio over four datasets: Twitter, Adult, US-GD and Non-Uni.

x. The empirical competitive ratio is evaluated for every baselines as well as our algorithm on top of every dataset, subject to various values of η_∞.

All our experiments are conducted on a laptop with Intel Core i7 CPU and 16GB memory. Since the algorithms are randomized, we repeat every run 10 times and take the average cost.

In Figure 1 we plot for every dataset a line-plot for all baselines and our algorithm, whose x-axis is η_∞ and y-axis is the empirical competitive ratio. From Figure 1, it can be seen that our algorithm performs consistently better than MEYERSON when η_∞ is relatively small, and it has comparable performance to MEYERSON when η_∞ is so large that the FOLLOW-PREDICT baseline loses control of the competitive ratio. For instance, in the Twitter dataset, our algorithm performs better than MEYERSON when $\eta_\infty \leq 2$, and when η_∞ becomes larger, our algorithm performs almost the same with MEYERSON while the ratio of FOLLOW-PREDICT baseline increases rapidly.

Furthermore, we observe that our performance lead over MERYERSON is especially significant on dataset Non-Uni whose opening cost is non-uniform. This suggests that our algorithm manages to use the predictions effectively in the non-uniform setting, even provided that the prediction is tricky to use since predictions at "good" locations could have large opening costs. In particular, our algorithm does a good job to find a "nearby" facility that has the correct opening cost and location tradeoff.

REFERENCES

Amir Ahmadi-Javid, Pardis Seyedi, and Siddhartha S. Syam. A survey of healthcare facility location. *Comput. Oper. Res.*, 79:223–263, 2017.

Antonios Antoniadis, Themis Gouleakis, Pieter Kleer, and Pavel Kolev. Secretary and online matching problems with machine learned advice. In *NeurIPS*, 2020.

Étienne Bamas, Andreas Maggiori, Lars Rohwedder, and Ola Svensson. Learning augmented energy minimization via speed scaling. In *NeurIPS*, 2020a.

Étienne Bamas, Andreas Maggiori, and Ola Svensson. The primal-dual method for learning augmented algorithms. In *NeurIPS*, 2020b.

Matej Cebecauer and L'uboš Buzna. Large-scale test data set for location problems. *Data in brief*, 17:267–274, 2018.

T. Hubert Chan, A. Guerqin, and M. Sozio. Twitter data set. URL https://github.com/fe6Bc5R4JvLkFkSeExHM/k-center.

Gabriella Divéki and Csanád Imreh. Online facility location with facility movements. *Central Eur. J. Oper. Res.*, 19(2):191–200, 2011.

Zvi Drezner and Horst W. Hamacher. *Facility location - applications and theory*. Springer, 2002.

Dheeru Dua and Casey Graff. UCI machine learning repository, 2017. URL https://archive.ics.uci.edu/ml/datasets/adult.

David Eisenstat, Claire Mathieu, and Nicolas Schabanel. Facility location in evolving metrics. In *ICALP*, volume 8573, pp. 459–470, 2014.

Dimitris Fotakis. On the competitive ratio for online facility location. *Algorithmica*, 50(1):1–57, 2008.

Dimitris Fotakis, Evangelia Gergatsouli, Themis Gouleakis, and Nikolas Patris. Learning augmented online facility location. *CoRR*, abs/2107.08277, 2021a.

Dimitris Fotakis, Loukas Kavouras, and Lydia Zakynthinou. Online facility location in evolving metrics. *Algorithms*, 14(3):73, 2021b.

Sreenivas Gollapudi and Debmalya Panigrahi. Online algorithms for rent-or-buy with expert advice. In *ICML*, volume 97 of *Proceedings of Machine Learning Research*, pp. 2319–2327. PMLR, 2019.

Zhihao Jiang, Pinyan Lu, Zhihao Gavin Tang, and Yuhao Zhang. Online selection problems against constrained adversary. In *ICML*, volume 139 of *Proceedings of Machine Learning Research*, pp. 5002–5012. PMLR, 2021.

Andreas Klose and Andreas Drexl. Facility location models for distribution system design. *Eur. J. Oper. Res.*, 162(1):4–29, 2005.

Gilbert Laporte, Stefan Nickel, and Francisco Saldanha da Gama. *Location science*. Springer, 2019.

Thodoris Lykouris and Sergei Vassilvitskii. Competitive caching with machine learned advice. *J. ACM*, 68(4):24:1–24:25, 2021.

Andres Muñoz Medina and Sergei Vassilvitskii. Revenue optimization with approximate bid predictions. In *NIPS*, pp. 1858–1866, 2017.

M. Teresa Melo, Stefan Nickel, and Francisco Saldanha-da-Gama. Facility location and supply chain management - A review. *Eur. J. Oper. Res.*, 196(2):401–412, 2009.

Ramgopal R. Mettu and C. Greg Plaxton. The online median problem. In *FOCS*, pp. 339–348. IEEE Computer Society, 2000.

Adam Meyerson. Online facility location. In *FOCS*, pp. 426–431. IEEE Computer Society, 2001.

Manish Purohit, Zoya Svitkina, and Ravi Kumar. Improving online algorithms via ML predictions. In *NeurIPS*, pp. 9684–9693, 2018.

Dhruv Rohatgi. Near-optimal bounds for online caching with machine learned advice. In *SODA*, pp. 1834–1845. SIAM, 2020.

Ryan A. Rossi and Nesreen K. Ahmed. The network data repository with interactive graph analytics and visualization. In *AAAI*, 2015. URL http://networkrepository.com.

Alexander Wei and Fred Zhang. Optimal robustness-consistency trade-offs for learning-augmented online algorithms. In *NeurIPS*, 2020.

A Proof of Lemma 3.9

Lemma A.1 (Restatement of 3.9)**.** *For every* $\ell \in [\underline{\ell}, \overline{\ell}]$,

$$\mathbb{E}\left[\sum_{i \in I_\ell} \mathrm{cost}_{\mathrm{M}}(x_i)\right] \le 18 \sum_{i \in I_\ell} d(x_i, f^*) + 32 w(f^*).$$

Proof. Recall that we relabeled and restricted the indices of points in X to $X(f^*) = \{x_1, \ldots, x_m\}$. However, in this proof, we also need to talk about the other points in the original data set (that do not belong to $X(f^*)$). Hence, to avoid confusions, we rewrite $X = \{y_1, \ldots, y_n\}$ (so $y_1, \ldots, y_n$ are the demand points in order), and we define $\sigma : [m] \to [n]$ that maps elements $x_j \in X(f^*)$ to its identity $y_i \in X$, i.e., $\sigma(j) = i$.

For $i \in [n]$, let

$$z_i^{\mathrm{M}} := \begin{cases} \hat{F}_{\mathrm{M}}(y_i) & \hat{F}_{\mathrm{M}}(y_i) \neq \emptyset \\ \text{none} & \text{otherwise} \end{cases},$$

and define z_i^{P} similarly. Since the Mey step opens at most one facility, z_i^{M} is either a singleton or "none". Finally, define $z_i := z_i^{\mathrm{M}} \cup z_i^{\mathrm{P}}$.

Define $\tau_\ell = \min\{i \in [n] \mid z_i^{\mathrm{M}} \neq \text{none and } d(f^*, z_i^{\mathrm{M}}) < 2^{\ell-1}\}$, we have

$$\mathbb{E}\left[\sum_{i \in I_\ell} \mathrm{cost}_{\mathrm{M}}(x_i) - 18 d(x_i, f^*)\right] = \mathbb{E}\left[\sum_{i \in [\tau_\ell]} \left(\mathrm{cost}_{\mathrm{M}}(y_i) - 18 d(y_i, f^*)\right) \mathbf{1}(i \in \sigma(I_\ell))\right]. \quad (1)$$

Next, we prove a stronger version by induction, and the induction hypothesis goes as follows. For every $j \in [n]$,

$$\mathbb{E}\left[\sum_{i=j}^{\tau_\ell} \left(\mathrm{cost}_{\mathrm{M}}(y_i) - 18 d(y_i, f^*)\right) \mathbf{1}(i \in \sigma(I_\ell)) \mid \mathcal{E}_{j-1}\right] \le 32 w(f^*),$$

where $\mathcal{E}_j = (z_1, \ldots, z_j)$. Clearly, applying this with $j = 1$ implies (1), hence it suffices to prove this hypothesis.

Base case. The base case is $j = n$, and we would do a backward induction on j. Since we condition on $\mathcal{E}_{j-1}$, the variables $\delta_0, \ldots, \delta_L$ in the Mey step of y_j, as well as wether or not $j \in \sigma(I_\ell)$, are fixed. Hence,

$$\mathbb{E}[\mathrm{cost}_{\mathrm{M}}(y_j) \mid \mathcal{E}_{j-1}] \le d(y_j, \bar{F}(y_j)) + \sum_{i=1}^{L} \frac{\delta_{i-1} - \delta_i}{2^i} \cdot 2^i \le 2 d(y_j, \bar{F}(y_j)).$$

This implies

$$\begin{aligned}
&\mathbb{E}\left[\sum_{i=j}^{\tau_\ell}\left(\mathrm{cost}_{\mathrm{M}}(y_i)-18d(y_i,f^*)\right)\mathbf{1}(i\in\sigma(I_\ell))\mid\mathcal{E}_{j-1}\right]\\
&\le \mathbb{E}[(\mathrm{cost}_{\mathrm{M}}(y_n)-18d(y_n,f^*))\mathbf{1}(n\in\sigma(I_\ell))\mid\mathcal{E}_{n-1}]\\
&\le 2d(y_n,\bar{F}(y_n))-18d(y_n,f^*)\mid_{n\in\sigma(I_\ell)}\\
&\le 2d(y_n,\bar{F}(y_n))-2d(y_n,f^*)\mid_{n\in\sigma(I_\ell)}\\
&\le 2d(\bar{F}(y_n),f^*)\mid_{n\in\sigma(I_\ell)}\\
&\le 2^{\ell+1}\le 2^{\ell^*+1}\le O(w(f^*)).
\end{aligned}$$

Inductive step. Next, assume the hypothesis holds for $j+1, j+2, \ldots, n$, and we would prove the hypothesis for j. We proceed with the following case analysis.

Case 1: $j \notin \sigma(I_\ell)$. We do a conditional expectation argument, and we have

$$\begin{aligned}
&\mathbb{E}\left[\sum_{i=j}^{\tau_\ell}\left(\mathrm{cost}_{\mathrm{M}}(y_i)-18d(y_i,f^*)\right)\mathbf{1}(i\in\sigma(I_\ell))\mid\mathcal{E}_{j-1}\right]\\
=&\sum_{z_j^{\mathrm{M}},z_j^{\mathrm{P}}}\Pr\left[z_j^{\mathrm{M}},z_j^{\mathrm{P}}\mid\mathcal{E}_{j-1}\right]\mathbb{E}\left[\sum_{i=j}^{\tau_\ell}\left(\mathrm{cost}_{\mathrm{M}}(y_i)-18d(y_i,f^*)\right)\mathbf{1}(i\in\sigma(I_\ell))\mid\mathcal{E}_{j-1},z_j^{\mathrm{M}},z_j^{\mathrm{P}}\right]\\
\le& 32w(f^*).
\end{aligned}$$

where the second step follows by induction hypothesis.

Case 2: $j \in I_\ell$ and $8d(y_j, f^*) > d(f^*, \bar{F}(y_j))$. We have

$$d(y_j,F(y_j))\le d(y_j,f^*)+d(f^*,F(y_j))\le 9d(y_j,f^*).$$

Hence $\mathbb{E}[\mathrm{cost}_{\mathrm{M}}(y_j)-18d(y_j,f^*)]\le 0$, and the hypothesis follows from a similar argument as in Case 1.

Case 3: $j \in I_\ell$, and $8d(y_j, f^*) \le d(f^*, \bar{F}(y_j))$. We have

$$\delta_{\ell^*}\le d(y_j,f^*)\le\frac{1}{8}d(f^*,\bar{F}(y_j))\le 2^{\ell-3}.$$

Let $D = \{f : d(f^*, f) \ge 2^{\ell-1}\} \cup \{\text{none}\}$. Let $k = \min\{k' \mid \delta_{k'} \le 2^{\ell-2}\}$. Then $\forall i \ge k$, $d(f_i, f^*) \le d(y_j, f_i) + d(y_j, f^*) < 2^{\ell-1}$.

$$\begin{aligned}
\Pr\left[\tau_\ell\le j\mid\mathcal{E}_{j-1}\right]&\ge\Pr\left[z_j^{\mathrm{M}}\notin D\mid\mathcal{E}_{j-1}\right]\ge\Pr\left[f_i\text{ is open for some }i\ge k\mid\mathcal{E}_{j-1}\right]\\
&=\min\left(\sum_{i\ge k}p_i,1\right)=\min\left(\sum_{i\ge k}\frac{\delta_{i-1}-\delta_i}{2^i},1\right)\\
&\ge\min\left(\frac{\delta_{k-1}-\delta_{\ell^*}}{2^{\ell^*}},1\right)\ge\min\left(\frac{2^{\ell-2}-2^{\ell-3}}{2^{\ell^*}},1\right)\ge\min\left(2^{\ell-3-\ell^*},1\right)\\
&\ge\frac{\delta_0}{16w(f^*)}\ge\frac{\mathbb{E}[\mathrm{cost}_{\mathrm{M}}(y_j)\mid\mathcal{E}_{j-1}]}{32w(f^*)},
\end{aligned}$$

where the second last inequality follows from $\delta_0 = d(y_j, \bar{F}(y_j)) \le 2^\ell$ and $w(f^*) = 2^{\ell^*-1}$. Thus,

$$\begin{aligned}
\text{LHS}=&\mathbb{E}[\mathrm{cost}_{\mathrm{M}}(y_j)\mid\mathcal{E}_{j-1}]\\
&+\sum_{t_j\in D}\sum_{z_j^{\mathrm{P}}}\Pr[z_j^{\mathrm{M}}=t_j,z_j^{\mathrm{P}}\mid\mathcal{E}_{j-1}]\,\mathbb{E}\left[\sum_{i=j+1}^{\tau_\ell}\left(\mathrm{cost}_{\mathrm{M}}(y_i)-18d(y_i,f^*)\right)\mathbf{1}(i\in\sigma(I_\ell))\mid\mathcal{E}_{j-1},z_j^{\mathrm{M}}=t_j,z_j^{\mathrm{P}}\right]\\
\le&\Pr\left[z_j^{\mathrm{M}}\notin D\mid\mathcal{E}_{j-1}\right]\cdot 32w(f^*)+\Pr\left[z_j^{\mathrm{M}}\in D\mid\mathcal{E}_{j-1}\right]\cdot 32w(f^*)\\
\le&\,32w(f^*),
\end{aligned}$$

where the second step follows by induction hypothesis. In summary, we have

$$\mathbb{E}\left[\sum_{i=1}^{\tau_\ell}\left(\mathrm{cost}_{\mathrm{M}}(y_i)-18d(y_i,f^*)\right)\mathbf{1}(i\in\sigma(I_\ell))\mid\mathcal{E}_{j-1}\right]\le 32w(f^*)$$

□

B PROOF OF LEMMA 3.5: WHEN $7\eta_\infty > 4w(f^*)$

The proof is mostly the same as in the other case which we prove in Section 3.1, and we only highlight the key differences. We use the same definition for parameters τ, $\underline{\ell}$, and $\overline{\ell}$. It can be verified that Lemma 3.7, 3.8 and 3.9 still holds and their proof in Section 3.1 still works.

However, Lemma 3.6 relies on that $7\eta_\infty \le 4w(f^*)$, which does not work in the current case. We provide Lemma B.2 in replacement of Lemma 3.6, which offers a slightly different bound but it still suffices for Lemma 3.5.

Lemma B.1. *For every $i<\tau$, $d(x_i,f^*)\ge w(f^*)$.*

Proof. We prove the statement by contradiction. Suppose $d(x_i,f^*)<w(f^*)$. Let $\delta_0,\delta_1,\dots,\delta_W$ be defined as in our algorithm. Then, we have

$$\delta_0=d(x_i,\bar{F}(x_i))\ge d(f^*,\bar{F}(x_i))-d(f^*,x_i)\ge d(f^*,\bar{F}(x_i))-w(f^*)>3w(f^*),$$

where the last inequality follows from the definition of τ that $d(f^*,\bar{F}(x_i))>4w(f^*)$. Let $k=\min\{k\mid\delta_k\le 3w(f^*)\}$. We have $k\ge 1$. We prove that a facility within a distance of $3w(f^*)$ from x_i must be open at this step.

$$\begin{aligned}\Pr\left[d(x_i,F(x_i))\le 3w(f^*)\right]&\ge\Pr\left[f_j\text{ is opened for some }j\ge k\right]\\&\ge\min\left(\sum_{j\ge k}p_k,1\right)\ge\min\left(\sum_{j=k}^{\ell^*}\frac{\delta_{j-1}-\delta_j}{2^j},1\right)\\&\ge\min\left(\frac{\delta_{k-1}-\delta_{\ell^*}}{2^{\ell^*}},1\right)\ge\min\left(\frac{3w(f^*)-w(f^*)}{2^{\ell^*}},1\right)=1,\end{aligned}$$

where the last inequality uses the fact that $\delta_{\ell^*}\le d(x_i,f^*)\le w(f^*)$. Consequently,

$$d(f^*,\hat{F}(x_{i+1}))\le d(f^*,F(x_i))\le d(x_i,F(x_i))+d(x_i,f^*)<4w(f^*),$$

which contradicts the definition of τ. □

Lemma B.2. $\mathbb{E}\left[\sum_{i\in[\tau]}\mathrm{cost}(x_i)\right]\le 6\,\mathbb{E}[\sum_{i\in[\tau]}d(x_i,f^*)]+4\cdot w(f^*)$.

Proof. On the arrival of x_i for each $i\in[\tau]$, let $\delta_0,\delta_1,\dots,\delta_L$ be defined as in our algorithm. We first study the connecting cost of request x_i. We have

$$d(x_i,F(x_i))\le\max\left(d(x_i,F(x_j)\backslash\hat{F}(x_i)),d(x_i,\hat{F}(x_i))\right)=\max\left(\delta_0,d(x_i,\hat{F}(x_i))\right).$$

We prove this value is at most $d(x_i,f^*)+2w(f^*)$. Suppose $\delta_0>d(x_i,f^*)+2w(f^*)$, as otherwise the statement holds. Recall that δ_j is non-increasing, let $k=\min\{k\mid\delta_k\le d(x_i,f^*)+2w(f^*)\}$. By the definition of k, we have that

$$\begin{aligned}\Pr\left[d(x_i,F(x_i))\le d(x_i,f^*)+2w(f^*)\right]&\ge\Pr[f_j\text{ is opened for some }j\ge k]\\&\ge\min\left(\sum_{j\ge k}p_j,1\right)\ge\min\left(\sum_{j\ge k}^{\ell^*}\frac{\delta_{j-1}-\delta_j}{2^j},1\right)\ge\min\left(\frac{\delta_{k-1}-\delta_{\ell^*}}{2^{\ell^*}},1\right)\\&\ge\min\left(\frac{d(x_i,f^*)+2w(f^*)-\delta_{\ell^*}}{2^{\ell^*}},1\right)\ge\min\left(\frac{w(f^*)}{2^{\ell^*-1}},1\right)=1.\end{aligned}$$

To sum up, we have shown that the connecting cost $d(x_i, F(x_i)) \leq d(x_i, f^*) + 2w(f^*)$.

Next, we study the expected opening cost $\mathbb{E}[w(\hat{F}(x_i))]$. We have

$$\mathbb{E}[w(\hat{F}(x_i))] = \sum_{j\in[L]} \Pr\left[\hat{F}(x_i) = \{f_j\}\right] \cdot w(f_j) \leq \sum_{j\in[L]} \min\left(\frac{\delta_{j-1}-\delta_j}{2^j}, 1\right) \cdot 2^{j-1}$$
$$\leq \sum_{j\leq \ell^*} 2^{j-1} + \sum_{j>\ell^*} \frac{\delta_{j-1}-\delta_j}{2^j} \cdot 2^{j-1} < 2^{\ell^*} + \frac{1}{2}\delta_{\ell^*} < 2w(f^*) + d(x_i, f^*).$$

By Lemma B.1, we conclude the proof of the statement:

$$\begin{aligned}
\mathbb{E}\left[\sum_{i\in[\tau]} \mathrm{cost}(x_i)\right] &= \mathbb{E}\left[\sum_{i\in[\tau]} \left(d(x_i, F(x_i)) + w(\hat{F}(x_i))\right)\right] \\
&\leq \mathbb{E}\left[\sum_{i\in[\tau]} (2d(x_i, f^*) + 4w(f^*))\right] \\
&\leq 6\,\mathbb{E}\left[\sum_{i<\tau} d(x_i, f^*) + (2d(x_\tau, f^*) + 4w(f^*))\right] \\
&\leq 6\,\mathbb{E}\left[\sum_{i\in[\tau]} d(x_i, f^*)\right] + 4w(f^*).
\end{aligned}$$

□

C Lower Bound

Without prediction, lower bound of the problem has already been studied by Fotakis (2008). Even when the facility cost is uniform and the metric space is a binary hierarchically well-separated tree (HST), no algorithm can achieve a competitive ratio better than $O(\frac{\log n}{\log\log n})$. We extend their construction to the setting with predictions, proving that when the predictions are not precise (i.e. $\eta_\infty > 0$), achieving a competitive ratio better than $\Omega(\frac{\log \frac{n\eta_\infty}{\mathrm{OPT}}}{\log\log n})$ is impossible.

Theorem C.1. *For every η_∞, there exists a class of inputs, such that every (randomized) online algorithm is $\Omega(\frac{\log \frac{n\eta_\infty}{\mathrm{OPT}}}{\log\log n})$-competitive, even when $\forall f \in \mathcal{F}, w(f) = 1$, and $\eta_1 = O(1)$.*

Proof. By Yao's principle, the expected cost of a randomized algorithm on the worst-case input is no better than the expected cost for a worst-case probability distribution on the inputs of the deterministic algorithm that performs best against that distribution. For each η_∞, we shall construct a family of randomized instance, so that the expected cost of any deterministic algorithm is at least $\Omega(\frac{\log \frac{n\eta_\infty}{\mathrm{OPT}}}{\log\log n})$-competitive.

We first import the construction for the classical online facility location problem by Fotakis (2008). Consider a hierarchically well-separated perfect binary tree. Let the distance between root and its children as D. For every vertex i of the tree, the distance between i and its children is $\frac{1}{m}$ times the distance between i and its parent. That is to say, the distance between a height i vertex and its children of height $i+1$ is $\frac{D}{m^i}$. Let the height of the tree be h.

The demand sequence is consisted of $h+1$ phases. In the i-th phase, m^{i-1} demand points arrive consecutively at some vertex of height i. The identity of the height i vertex is independently and uniformly chosen between the two children of $(i-1)$-th phase vertex.

Consider the solution that opens only one facility at the leaf node that is the h-th phase demand vertex and assign all demand points to the leaf. The cost of this solution equals

$$1 + \sum_{i=0}^{h-1} m^i \sum_{j=i}^{h-1} \frac{D}{m^j} \leq \sum_{i=0}^{h-1} m^i \frac{D}{m^i} \frac{m}{m-1} = 1 + hD\frac{m}{m-1}.$$

This value serves as an upper bound of the optimal cost.

Next, we describe the predictions associated with the demand points. Intuitively, we try the best to hide the identity of the leaf node in the last phase. We denote the leaf node as f^*, which is also the open facility in the above described solution. Our prediction is produced according to the following rule. When the distance between the current demand point and f^* is less than η_∞, let the prediction be the same as the demand vertex. Otherwise, let the prediction between the vertex on the path from the current demand point to f^*, whose distance to f^* equals η_∞.

For any deterministic algorithm, we consider the lower bound of its cost. The whole process contains $h+1$ phases, so we first overlook many phases until the subtree of the current demand vertex has diameter less than η_∞. Let it be the h'-th phase. We now focus on the cost induced after the h'-th phase, the predictions are useless as they are just the demand points. When the m^i demand points at vertex of height i comes, we consider the following two cases:

Case 1: There is no facility opened in the current subtree. Then we either open a new facility bearing an opening cost 1 or assign the current demands to facilities outside the subtree bearing a large connection cost at least $\frac{D}{m^{i-1}}$ per demand. So the cost of this vertex is $\min\{1, Dm\}$.

Case 2: There has been at least one facility opened in the current subtree. Since the next vertex is uniformly randomly chosen in its two children, the expected number of facility that will not enter the next phase subtree is at least$\frac{1}{2}$. We call it abandoned. The expected sum of abandoned facility is $\frac{1}{2}$ of the occurrence of case 2. Thus the expected cost in every occurrence of case 2 is $\frac{1}{2}$.

We shall carefully choose D, m, h so that 1) $mD = 1$; 2) the total number of demand points is n, i.e. $\sum_{i=0}^{h} m^i = n$; and 3)$\frac{D}{m^{h'}} = \eta_\infty$. These conditions give that $(h+1)\log m = \log n$, $(h'+1)\log m = -\log \eta_\infty$. Consequently, $h - h' = \frac{\log n\eta_\infty}{\log m}$.

To sum up, an lower bound of any deterministic algorithm is $(h-h')\min\{\frac{1}{2}, mD\} = \Omega\left(\frac{\log n\eta_\infty}{\log m}\right)$, while the optimal cost is at most $1 + hD\frac{m}{m-1} = O\left(\frac{\log n}{m\log m}\right)$. Setting $m = \frac{\log n}{\log\log n}$, we get the lower bound of $\Omega(\frac{\log \frac{n\eta_\infty}{\mathrm{OPT}}}{\log\log n})$. It is straightforward to check that the summation of the prediction error $\eta_1 = O(1)$ in the above construction and it is easy to see that it cannot help with a larger η_1, since we can always provide a prediction that is farther from f^* than the demand point itself. $\square$