

## MIT Open Access Articles

*Top-k Ranking Bayesian Optimization*

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

**Citation:** Nguyen, Quoc Phong, Tay, Sebastian, Low, Bryan Kian Hsiang, Jaillet, Patrick and Intelligence, Assoc Advancement Artificial. 2021. "Top-k Ranking Bayesian Optimization." THIRTY-FIFTH AAAI CONFERENCE ON ARTIFICIAL INTELLIGENCE, THIRTY-THIRD CONFERENCE ON INNOVATIVE APPLICATIONS OF ARTIFICIAL INTELLIGENCE AND THE ELEVENTH SYMPOSIUM ON EDUCATIONAL ADVANCES IN ARTIFICIAL INTELLIGENCE, 35.

**Published Version:** <https://ojs.aaai.org/index.php/AAAI/article/view/17103>

**Publisher:**

**Permanent Link:** <https://hdl.handle.net/1721.1/143694>

**Version:** Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

**Terms of use:** Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.



# Top- $k$ Ranking Bayesian Optimization

Quoc Phong Nguyen,<sup>1</sup> Sebastian Tay,<sup>1</sup> Bryan Kian Hsiang Low,<sup>1</sup> Patrick Jaillet<sup>2</sup>

<sup>1</sup>Dept. of Computer Science, National University of Singapore, Republic of Singapore

<sup>2</sup>Dept. of Electrical Engineering and Computer Science, MIT, USA

qphong@comp.nus.edu.sg, sebastian.tay@u.nus.edu, lowkh@comp.nus.edu.sg, jaillet@mit.edu

## Abstract

This paper presents a novel approach to *top-k ranking Bayesian optimization* (top- $k$  ranking BO) which is a practical and significant generalization of preferential BO to handle top- $k$  ranking and tie/indifference observations. We first design a surrogate model that is not only capable of catering to the above observations, but is also supported by a classic random utility model. Another equally important contribution is the introduction of the first information-theoretic acquisition function in BO with preferential observation called *multinomial predictive entropy search* (MPES) which is flexible in handling these observations and optimized for all inputs of a query jointly. MPES possesses superior performance compared with existing acquisition functions that select the inputs of a query one at a time greedily. We empirically evaluate the performance of MPES using several synthetic benchmark functions, CIFAR-10 dataset, and SUSHI preference dataset.

## 1 Introduction

*Bayesian optimization* (BO) is an efficient approach to optimize expensive-to-evaluate black-box objective functions (i.e., possibly noisy, non-convex, and/or with no closed-form expression/derivative) (Brochu, Cora, and de Freitas 2010). In practice, direct access to function evaluations may not be always possible (González et al. 2017). For example, a diner tasting two different dishes (i.e., inputs) can easily tell which dish he/she prefers, but it is relatively difficult for the diner to articulate a rigorous numeric value representing the taste of each dish (i.e., a function). The difficulty in providing such a value arises from the need to taste all possible dishes (i.e., inputs) and assess the difference in the flavors of these dishes. On the other hand, specifying the preference between dishes (i.e., inputs) is so natural that it becomes our daily dining routine. Also, an inherent property of our preference, which should be built into the model, is our inability to elucidate the preference between very similar choices (i.e., indifference or a tie). For example, it is hardly possible for us to tell the difference between a cup of coffee with 20% of sugar and another with 21% of sugar.

To boost the practicality of BO, recent works on *preferential BO* (Dewancker, Bauer, and McCourt 2017; González

et al. 2017) have attempted to replace direct (but noisy) function evaluations/values with noisy preferences between inputs. In particular, given two inputs, the higher the objective function value at an input is, the more likely it is preferred. The observation in these works is limited to a preference between a pair of inputs, i.e., a *pairwise preference* (Dewancker, Bauer, and McCourt 2017; González et al. 2017). Furthermore, inputs in the pair are searched one after the other. Ideally, we would like to search for the input pair jointly as the function values at these inputs are correlated.

Regarding the model, the work of González et al. (2017) directly applies a *Gaussian process* (GP) to model a latent preference function whose input is a pair of the objective function’s inputs. This model suffers from 3 disadvantages: The input dimension of the GP is twice that of the objective function, the objective function is not modeled directly, and ties are not modeled. The second drawback leads to the use of a computationally expensive soft-Copeland score to estimate the maximizer of the objective function. On the other hand, the work of Dewancker, Bauer, and McCourt (2017) employs the generalized Bradley-Terry model. Yet, it suffers from a crude mean field approximation which implies that the posterior beliefs of the function values at different inputs are independent from one another. Note that there are several works with preference-based observations in bandit literature (Busa-Fekete, Hüllermeier, and Mesaoudi-Paul 2018). The most relevant work to preferential BO is the work of Sui et al. (2017) where there is an infinite number of dependent arms modeled with a GP. However, it does not have a regret analysis like the other bandit algorithms. Besides, a probability density is modeled with a GP which allows negative values.

This paper presents an approach that can resolve both existing issues on the model and the acquisition function. Our model is inspired from the multinomial logit model and its generalization to rankings and ties. Combining with a GP, our model can be interpreted as a GP regression model with an i.i.d. Gumbel noise. As the GP directly models the underlying objective function, the maximizer of the objective function can be estimated with the maximizer of the GP posterior mean function like in the conventional BO, which is less computationally intensive than the soft-Copeland score in (González et al. 2017). Our model is capable of handling the observation as a ranking of the top- $k$  inputs in a finite set of inputs, i.e., *top-k rankings* (Sec. 2.1), and the possibility

of a tie/indifference in the observation (Sec. 2.2). The former subsumes the pairwise preference (i.e., a top-1 ranking of 2 inputs) in the existing works (Dewancker, Bauer, and McCourt 2017; González et al. 2017). We call this generalized problem the *top-k ranking Bayesian optimization* (top- $k$  BO). Although our GP model has a non-Gaussian likelihood, it can be trained with *variational inference* (Sec. 3).

While information-theoretic acquisition functions (Hennig and Schuler 2012; Hernández-Lobato, Hoffman, and Ghahramani 2014; Wang and Jegelka 2017) have been investigated extensively in the conventional BO and demonstrated promising performance, such a principled acquisition function, to the best of our knowledge, has not been explored in BO with preferential observation. Therefore, to efficiently exploit the posterior belief provided by our model, we derive the first information-theoretic acquisition function for BO with preferential observation that is capable of handling different types of observation introduced in this paper. It maximizes the information gain on the maximizer of the objective function through observing the top- $k$  ranking observation, which we call *multinomial predictive entropy search* (MPES) (Sec. 4). Apart from the fact that MPES is rooted in information theory, it can jointly search for all inputs of the query, which differs from existing acquisition functions for preferential BO (Dewancker, Bauer, and McCourt 2017; González et al. 2017) that search for inputs of a query one at a time greedily. We empirically evaluate our model and our acquisition function with different types of observation using several synthetic benchmark functions, CIFAR-10 dataset, and SUSHI preference dataset (Sec. 5).

## 2 Multinomial Logit Model and Top- $k$ Ranking Observations

Let  $f : \mathcal{X} \rightarrow \mathbb{R}$  be an unknown objective function defined on a bounded input domain  $\mathcal{X} \subset \mathbb{R}^d$ . A function evaluation/value at an input  $\mathbf{x} \in \mathcal{X}$  is denoted as  $f_{\mathbf{x}} \in \mathbb{R}$ . The goal is to search for the maximizer  $\operatorname{argmax}_{\mathbf{x} \in \mathcal{X}} f_{\mathbf{x}}$  by observing preferences between different inputs, i.e., observing the choice between different inputs based on noisy evaluation of  $f$  at these inputs. The noisy evaluation of  $f$  can be viewed as the *utility function*  $u$  decomposed into  $u_{\mathbf{x}} \triangleq f_{\mathbf{x}} + \epsilon_{\mathbf{x}}$  where  $\epsilon_{\mathbf{x}}$  is a random noise. This noise represents unknown factors that affect the preference but are not captured in our objective function  $f_{\mathbf{x}}$ , which is a practical consideration. Recall our dining example in Sec. 1, the objective function may not capture the effect of the food temperature or the diner’s hunger on the preference of the dish. Following the *random utility model* (Marschak 1959), an input  $\mathbf{x}$  is preferred over another input  $\mathbf{x}'$  (i.e., denoted as  $\mathbf{x} \succ \mathbf{x}'$ ) if the difference  $u_{\mathbf{x}} - u_{\mathbf{x}'}$  in the unknown utility function values at  $\mathbf{x}$  and  $\mathbf{x}'$  is at least a threshold  $\delta \geq 0$ . In other words,  $p(\mathbf{x} \succ \mathbf{x}') = p(u_{\mathbf{x}} - u_{\mathbf{x}'} \geq \delta)$ . The threshold  $\delta$  enables the possibility of a tie between 2 inputs, which reflects the real-world scenario where one is indifferent between  $\mathbf{x}$  and  $\mathbf{x}'$  due to their similar utility values, i.e.,  $u_{\mathbf{x}}$  and  $u_{\mathbf{x}'}$  are not sufficiently far apart. It subsumes an extreme case of  $\delta = 0$ , i.e., there is no tie in the observation such as in (González et al. 2017).

To specify the noise  $\epsilon_{\mathbf{x}}$ , a common approach in the litera-

ture of preference learning with GP (Chu and Ghahramani 2005; González et al. 2017) is to assume a Gaussian noise. However, it is difficult to extend such a model of pairwise preferences to rankings, as explained in (Nguyen et al. 2020). On the other hand, we present a refreshing approach of modeling  $\epsilon_{\mathbf{x}}$  as a Gumbel noise such that we can leverage the well-established multinomial logit model (McFadden 1974) to enable rankings and ties in BO.

Under the Gumbel noise and given the objective function values, the probability that an input  $\mathbf{x}$  is preferred over a finite set  $\mathcal{C} \subset \mathcal{X}$  of inputs (i.e., may or may not include  $\mathbf{x}$ ) has a closed-form expression, as derived in (Nguyen et al. 2020):

$$p(\mathbf{x} \succ \mathcal{C} \setminus \{\mathbf{x}\} | \mathbf{f}_{\mathcal{C} \cup \{\mathbf{x}\}}; \delta) = \frac{e^{f_{\mathbf{x}}}}{e^{f_{\mathbf{x}}} + \sum_{\mathbf{x}' \in \mathcal{C} \setminus \{\mathbf{x}\}} e^{f_{\mathbf{x}'} + \delta}} \quad (1)$$

where  $p(\mathbf{x} \succ \mathcal{C} \setminus \{\mathbf{x}\}) \triangleq p(\forall \mathbf{x}' \in \mathcal{C} \setminus \{\mathbf{x}\} \quad \mathbf{x} \succ \mathbf{x}')$  and  $\mathbf{f}_{\mathcal{C} \cup \{\mathbf{x}\}} \triangleq (f_{\mathbf{x}''})_{\mathbf{x}'' \in \mathcal{C} \cup \{\mathbf{x}\}}$  consists of function values at inputs in  $\mathcal{C} \cup \{\mathbf{x}\}$  (Cantillo, Amaya, and Ortúzar 2010). As a result, when  $\delta = 0$ , inputs with equal objective function values have equal probabilities of being preferred while inputs with higher objective function values are exponentially more likely to be preferred. When there is no tie (i.e.,  $\delta = 0$ ), the model can be generalized to accept a ranking (i.e.,  $\mathbf{x}_1 \succ \mathbf{x}_2 \succ \dots \succ \mathbf{x}_m$ ) as an observation following the Plackett-Luce model (Luce 1959; Plackett 1975):

$$\begin{aligned} & p(\mathbf{x}_1 \succ \mathbf{x}_2 \succ \dots \succ \mathbf{x}_m | \mathbf{f}_{\cup_{i=1}^m \{\mathbf{x}_i\}}) \\ &= \prod_{i=1}^{m-1} p(\mathbf{x}_i \succ \cup_{j=i+1}^m \{\mathbf{x}_j\} | \mathbf{f}_{\cup_{j=i}^m \{\mathbf{x}_j\}}; \delta = 0). \end{aligned} \quad (2)$$

In the following subsections, we introduce the notations for 2 different types of observation: the top- $k$  ranking and the top-1 ranking with ties.

Apart from the difference in the noise model, the work of González et al. (2017), which defines the probability that an input  $\mathbf{x}$  is preferred over another input  $\mathbf{x}'$  as  $p(\mathbf{x} \succ \mathbf{x}') = (1 + \exp(f_{\mathbf{x}'} - f_{\mathbf{x}}))^{-1}$ , can be viewed as a special case of (1). Therefore, by interpreting the formulation from the multinomial logit model with ties, our model is a generalization of the model in (González et al. 2017).

### 2.1 Top- $k$ Ranking

When there is no tie ( $\delta = 0$ ), let a *top- $k$  ranking* (i.e., a ranking of the top- $k$  inputs in a finite set) over a finite set  $\mathcal{C} \subset \mathcal{X}$  (where  $0 < k < |\mathcal{C}|$ ) of inputs (i.e., interpreted as choices) be denoted as  $\mathbf{o}_{\mathcal{C}}^k \triangleq \{\mathbf{o}_{\mathcal{C}}^k(i)\}_{i=1}^k$  which is an ordered set of  $k$  inputs in descending order of preference in  $\mathcal{C}$ , that is,  $\mathbf{o}_{\mathcal{C}}^k(i) \in \mathcal{C}$  and  $\mathbf{o}_{\mathcal{C}}^k(i) \succ \mathbf{o}_{\mathcal{C}}^k(j)$  for all  $i < j$ . For example,  $\mathbf{o}_{\mathcal{C}}^k(1)$  is the most preferred input in  $\mathcal{C}$ , and similarly,  $\mathbf{o}_{\mathcal{C}}^k(i)$  is the  $i$ -th most preferred input in  $\mathcal{C}$ . From (2), the probability of a top- $k$  ranking is expressed as

$$p(\mathbf{o}_{\mathcal{C}}^k | \mathbf{f}_{\mathcal{C}}) = \prod_{i=1}^k p(\mathbf{o}_{\mathcal{C}}^k(i) \succ \mathcal{C} \setminus \cup_{j=1}^i \{\mathbf{o}_{\mathcal{C}}^k(j)\} | \mathbf{f}_{\mathcal{C}}; \delta = 0). \quad (3)$$

When  $|\mathcal{C}| = k + 1 = 2$ , a top- $k$  ranking reduces to a pairwise preference (i.e., between a pair of inputs) which is the observation considered in (González et al. 2017).

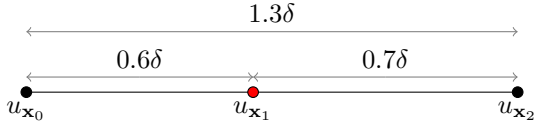


Figure 1: A counter-example based on the transitivity property of a tie relation between a pair of inputs.

Note that  $\mathbf{o}_C^{|C|-1}$  is a (full) ranking of all inputs in  $\mathcal{C}$ . Its probability (i.e., specified in (3) when  $k = |C| - 1$ ) differs from that of a batch of pairwise preferences, the latter of which is the product of probabilities of pairwise preferences in the batch. This is different from the work of González et al. (2017) which claims that rankings can be trivially mapped to pairwise preferences. In fact, simple approaches of mapping rankings to pairwise preferences can violate a probability axiom, as shown in (Nguyen et al. 2020).

## 2.2 Top-1 Ranking with Ties

In practice, we are often incapable of stating a strict preference between inputs with similar utility function values. One may ignore the observation in this case if the model can only handle strict preference. However, overlooking this observation reduces the query efficiency of the BO algorithm. Therefore, apart from the strict preference between inputs, tie/indifference should be allowed in the model to capture this observation.

We investigate one such possibility of ties (i.e.,  $\delta > 0$ ) in the observation when  $k = 1$  (i.e., the top-1 ranking with ties). To simplify notation, we denote  $\mathbf{o}_C$  as the only input in  $\mathbf{o}_C^1$ , i.e.,  $\mathbf{o}_C$  is the most preferred input in  $\mathcal{C}$ . Let  $\mathbf{o}_C = \emptyset$  denote the event where there exists a tie in finding the most preferred input in  $\mathcal{C}$ . The probability of a tie can be expressed as follows:

$$p(\mathbf{o}_C = \emptyset | \mathbf{f}_C; \delta) = 1 - \sum_{\mathbf{o}_C \in \mathcal{C}} p(\mathbf{o}_C | \mathbf{f}_C; \delta) \quad (4)$$

s.t.  $p(\mathbf{o}_C | \mathbf{f}_C; \delta) = p(\mathbf{o}_C \succ \mathcal{C} \setminus \{\mathbf{o}_C\} | \mathbf{f}_C; \delta)$  is specified in (1).

**Remark 1** When  $k > 1$ , the probability of a tie observation cannot be computed as straightforwardly as (4) because the tie relation between a pair of inputs (denoted as  $\mathbf{x} \sim \mathbf{x}'$ ) is not transitive. Fig. 1 shows a counter-example where  $\mathbf{x}_0 \sim \mathbf{x}_1$  ( $|u_{\mathbf{x}_0} - u_{\mathbf{x}_1}| < \delta$ ) and  $\mathbf{x}_1 \sim \mathbf{x}_2$  ( $|u_{\mathbf{x}_1} - u_{\mathbf{x}_2}| < \delta$ ) do not lead to  $\mathbf{x}_0 \sim \mathbf{x}_2$  ( $|u_{\mathbf{x}_0} - u_{\mathbf{x}_2}| < \delta$ ). This is further elaborated in (Nguyen et al. 2020) and therefore left for future work.

## 3 Variational Inference with Top- $k$ Ranking Observations

We employ a noiseless *Gaussian process* (GP) to model the unknown objective function  $f$ . In other words, function values at every finite subset of  $\mathcal{X}$  follow a multivariate Gaussian distribution (Rasmussen and Williams 2006) whose *prior* distribution is specified by the GP prior mean and covariance  $k_{\mathbf{x}, \mathbf{x}'} \triangleq \text{cov}[f_{\mathbf{x}}, f_{\mathbf{x}'}]$ . As the preference probability in (1) does not change when  $f$  is shifted by a constant (i.e., shift-invariant), the GP prior mean is set to 0. The covariance is defined by the *squared exponential* kernel, i.e.,

$k_{\mathbf{x}, \mathbf{x}'} \triangleq \sigma_s^2 \exp(-0.5(\mathbf{x} - \mathbf{x}')^\top \Lambda^{-2}(\mathbf{x} - \mathbf{x}'))$  where the hyperparameters consist of the length-scales  $\Lambda \triangleq \text{diag}[l_1, \dots, l_d]$  and the signal variance  $\sigma_s^2$ .

Unlike the work of González et al. (2017) that models the function  $f_{\mathbf{x}'} - f_{\mathbf{x}}$  (i.e., a function with the input as a pair  $[\mathbf{x}, \mathbf{x}']$ ), we directly model the objective function  $f_{\mathbf{x}}$  which requires only half the dimension of the input of  $f_{\mathbf{x}'} - f_{\mathbf{x}}$ . Therefore, while our GP can be intuitively interpreted as the belief over the unknown objective function, the GP in (González et al. 2017) cannot. As a result, in our model, the maximizer of the mean function of the posterior GP belief can be viewed as an estimate of the maximizer of the objective function, while the work of González et al. (2017) needs to introduce the soft-Copeland score to estimate the maximizer of the objective function. Evaluating the soft-Copeland score requires the use of Monte-Carlo integration over  $\mathbf{x} \in \mathcal{X}$ , which is prohibitively expensive for problems with high input dimension. Another advantage of our model is the ability of handling ranking observation, which is difficult to extend from the GP model of pairwise preferences in (González et al. 2017), as explained in (Nguyen et al. 2020).

Let  $\mathcal{D}$  denote the observations (e.g., top- $k$  rankings and top-1 rankings with ties) in a BO iteration,  $\mathcal{X}_{\mathcal{D}}$  denote the set of distinct inputs in  $\mathcal{D}$ , and  $\mathbf{f}_{\mathcal{X}_{\mathcal{D}}}$  denote the function values evaluated at  $\mathcal{X}_{\mathcal{D}}$ . The likelihood  $p(\mathcal{D} | \mathbf{f}_{\mathcal{X}_{\mathcal{D}}})$  of the observations is specified in Sec. 2 and is not a Gaussian distribution. Hence, the GP posterior belief given the observations does not have a closed-form expression, unlike that of GP regression in conventional BO. To estimate the posterior belief, we use the variational inference technique to learn an approximate Gaussian posterior belief  $q(\mathbf{f}_{\mathcal{X}_{\mathcal{D}}}) \triangleq \mathcal{N}(\mathbf{f}_{\mathcal{X}_{\mathcal{D}}} | \boldsymbol{\mu}_{\mathcal{X}_{\mathcal{D}}}, \boldsymbol{\Sigma}_{\mathcal{X}_{\mathcal{D}}})$  of  $\mathbf{f}_{\mathcal{X}_{\mathcal{D}}}$  given  $\mathcal{D}$  by maximizing the *evidence lower bound* (ELBO):<sup>1</sup>

$$\int q(\mathbf{f}_{\mathcal{X}_{\mathcal{D}}}) \log p(\mathcal{D} | \mathbf{f}_{\mathcal{X}_{\mathcal{D}}}) \, d\mathbf{f}_{\mathcal{X}_{\mathcal{D}}} - \int q(\mathbf{f}_{\mathcal{X}_{\mathcal{D}}}) \log \frac{q(\mathbf{f}_{\mathcal{X}_{\mathcal{D}}})}{p(\mathbf{f}_{\mathcal{X}_{\mathcal{D}}})} \, d\mathbf{f}_{\mathcal{X}_{\mathcal{D}}}. \quad (5)$$

The ELBO can be maximized with a stochastic gradient ascent algorithm to obtain  $\boldsymbol{\mu}_{\mathcal{X}_{\mathcal{D}}}, \boldsymbol{\Sigma}_{\mathcal{X}_{\mathcal{D}}}$ , the GP hyperparameters, and the threshold  $\delta$  if ties exist, i.e., by drawing a random mini-batch of  $\mathbf{f}_{\mathcal{X}_{\mathcal{D}}}$  from the current variational distribution  $q(\mathbf{f}_{\mathcal{X}_{\mathcal{D}}})$  in each iteration to optimize (5).<sup>2</sup> Since our GP directly models the objective function, the length-scales of the GP can be interpreted as the rate of decay of the spatial correlation of the objective function in terms of the Euclidean distance between the inputs. Given that the approximate posterior belief  $q(\mathbf{f}_{\mathcal{X}_{\mathcal{D}}})$  is a multivariate Gaussian distribution  $\mathcal{N}(\boldsymbol{\mu}_{\mathcal{X}_{\mathcal{D}}}, \boldsymbol{\Sigma}_{\mathcal{X}_{\mathcal{D}}})$ , the posterior predictive belief of the function values at any finite subset  $\mathcal{X}' \subset \mathcal{X}$  given  $\mathcal{D}$  (i.e., by marginalizing over  $\mathbf{f}_{\mathcal{X}_{\mathcal{D}}}$ ) is also a multivariate Gaussian distribution whose mean and covariance are specified as follows:

$$\begin{aligned} \boldsymbol{\mu}_{\mathcal{X}'} &\triangleq \mathbf{K}_{\mathcal{X}', \mathcal{X}_{\mathcal{D}}} \mathbf{K}_{\mathcal{X}_{\mathcal{D}}, \mathcal{X}_{\mathcal{D}}}^{-1} \boldsymbol{\mu}_{\mathcal{X}_{\mathcal{D}}} \\ \boldsymbol{\Sigma}_{\mathcal{X}'} &\triangleq \mathbf{K}_{\mathcal{X}', \mathcal{X}'} - \mathbf{K}_{\mathcal{X}', \mathcal{X}_{\mathcal{D}}} \boldsymbol{\Lambda} \mathbf{K}_{\mathcal{X}_{\mathcal{D}}, \mathcal{X}'} \end{aligned} \quad (6)$$

<sup>1</sup>Like the definition of  $\mathbf{f}_{\mathcal{X}_{\mathcal{D}}}, \boldsymbol{\mu}_{\mathcal{X}_{\mathcal{D}}}$  is a vector of posterior mean values at  $\mathcal{X}_{\mathcal{D}}$  and  $\boldsymbol{\Sigma}_{\mathcal{X}_{\mathcal{D}}}$  is a covariance matrix whose elements are the posterior covariance between function values at  $\mathcal{X}_{\mathcal{D}}$ .

<sup>2</sup>To ensure  $\boldsymbol{\Sigma}_{\mathcal{X}_{\mathcal{D}}}$  is a positive-definite matrix, we optimize its square root lower-triangular matrix. A regularizer over the length-scales can be applied to ensure sufficiently large length-scales.

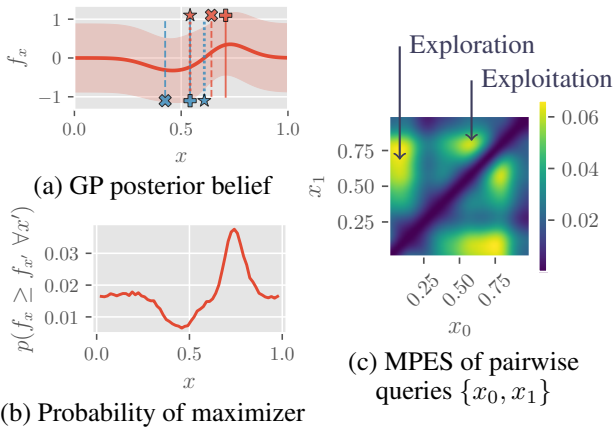


Figure 2: An example of top- $k$  BO with 3 pairwise preferences: Inputs of each pair are represented as the  $x$  values of cross, star, and plus markers. The red markers are the preferred inputs (i.e., plotted higher).

where  $\mathbf{K}_{\mathcal{X}_D, \mathcal{X}_D} \triangleq (k_{\mathbf{x}\mathbf{x}'}),_{\mathbf{x}, \mathbf{x}' \in \mathcal{X}_D}$ ,  $\mathbf{K}_{\mathcal{X}' \mathcal{X}'} \triangleq (k_{\mathbf{x}\mathbf{x}'}),_{\mathbf{x}, \mathbf{x}' \in \mathcal{X}'}$ ,  $\mathbf{K}_{\mathcal{X}' \mathcal{X}_D} \triangleq (k_{\mathbf{x}\mathbf{x}'}),_{\mathbf{x} \in \mathcal{X}', \mathbf{x}' \in \mathcal{X}_D}$ ,  $\mathbf{K}_{\mathcal{X}_D \mathcal{X}'}$   $\triangleq \mathbf{K}_{\mathcal{X}' \mathcal{X}_D}^\top$ , and  $\mathbf{\Lambda} \triangleq \mathbf{K}_{\mathcal{X}_D \mathcal{X}_D}^{-1} (\mathbf{K}_{\mathcal{X}_D \mathcal{X}_D} - \mathbf{\Sigma}_{\mathcal{X}_D}) \mathbf{K}_{\mathcal{X}_D \mathcal{X}_D}^{-1}$ . Note that sparse GP models can be used to reduce the time complexity of  $\mathcal{O}(|\mathcal{X}_D|^3)$  (i.e., arising from the matrix inversion  $\mathbf{K}_{\mathcal{X}_D \mathcal{X}_D}^{-1}$ ) (Quiñero-Candela and Rasmussen 2005), but the full GP model is used in this paper to be comparable with the models used by other baseline methods.

Fig. 2a shows the GP posterior belief given the observations consisting of 3 pairwise preferences. Due to noise, there is an incorrect preference between inputs plotted as stars, i.e., the input with the smaller  $f_x$  is observed as being preferred. It can be observed that the difference in the function values at the inputs in the incorrect preference (plotted as stars) is smaller than that at the inputs in the correct preferences (crosses and pluses), which aligns with our formulation in (1). Given the GP posterior belief, one can use the maximizer of the GP posterior mean function as an estimate of the maximizer of the objective function. In the next section, we utilize this GP posterior belief to design an information-theoretic acquisition function that can efficiently guide the query selection to search for the maximizer of the objective function.

## 4 Multinomial Predictive Entropy Search (MPES)

While information-theoretic acquisition functions have been explored extensively in conventional BO (Hennig and Schuler 2012; Hernández-Lobato, Hoffman, and Ghahramani 2014; Ru et al. 2018; Shah and Ghahramani 2015; Wang and Jegelka 2017), they have not been investigated for preferential BO or our generalized top- $k$  BO. Therefore, we propose to construct a principled acquisition function based on information theory to select the next query (i.e., a set  $\mathcal{C}$  of inputs), which we call *multinomial predictive entropy search* (MPES). Specifically, the next query is selected such that it maximizes the information gain on the maximizer  $\mathbf{x}_*$  of the objective

function through observing the (top- $k$  ranking) observation at the query. Let  $\mathbf{o}_C^k$  denote the observation given a query  $\mathcal{C}$ . The observation can be either a top- $k$  ranking or a top-1 ranking with ties (Sec. 2). The information gain is measured by the mutual information between  $\mathbf{x}_*$  and  $\mathbf{o}_C^k$ , which is interpreted as the reduction in the entropy (uncertainty) of the maximizer  $\mathbf{x}_*$  given the observation  $\mathbf{o}_C^k$ :

$$I(\mathbf{o}_C^k; \mathbf{x}_* | \mathcal{D}) \triangleq H(p(\mathbf{x}_* | \mathcal{D})) - \mathbb{E}_{p(\mathbf{o}_C^k | \mathcal{D})} [H(p(\mathbf{x}_* | \mathcal{D}, \mathbf{o}_C^k))]$$

where  $H(p(\mathbf{x}_* | \mathcal{D})) \triangleq - \int_{\mathbf{x}_*} p(\mathbf{x}_* | \mathcal{D}) \log p(\mathbf{x}_* | \mathcal{D}) d\mathbf{x}_*$  denotes the entropy of  $\mathbf{x}_*$  and  $\mathbb{E}_{p(\mathbf{o}_C^k | \mathcal{D})} [H(p(\mathbf{x}_* | \mathcal{D}, \mathbf{o}_C^k))]$  denotes the conditional entropy of  $\mathbf{x}_*$  given  $\mathbf{o}_C^k$ . However, the above expression requires a prohibitively expensive evaluation of  $p(\mathbf{x}_* | \mathcal{D}, \mathbf{o}_C^k)$  for all possible values of  $\mathbf{o}_C^k$ . This issue also exists in several information-theoretic acquisition functions in conventional BO such as in (Hennig and Schuler 2012; Villemonteix, Vazquez, and Walter 2009). Therefore, we employ the symmetric property of mutual information to express the acquisition function as

$$\begin{aligned} I(\mathbf{o}_C^k; \mathbf{x}_* | \mathcal{D}) &= H(p(\mathbf{o}_C^k | \mathcal{D})) - \mathbb{E}_{p(\mathbf{x}_* | \mathcal{D})} [H(p(\mathbf{o}_C^k | \mathcal{D}, \mathbf{x}_*))] \\ &= \sum_{\mathbf{o}_C^k} \int_{\mathbf{x}_*} p(\mathbf{o}_C^k, \mathbf{x}_* | \mathcal{D}) \log \frac{p(\mathbf{o}_C^k | \mathcal{D}, \mathbf{x}_*)}{p(\mathbf{o}_C^k | \mathcal{D})} d\mathbf{x}_*. \end{aligned} \quad (7)$$

The next query is selected as  $\arg\max_{\mathcal{C} \in \mathcal{X}^{|\mathcal{C}|}} I(\mathbf{o}_C^k; \mathbf{x}_* | \mathcal{D})$ , which trades off between exploration (i.e., maximizing  $H(p(\mathbf{o}_C^k | \mathcal{D}))$ ) and exploitation (i.e., minimizing  $\mathbb{E}_{p(\mathbf{x}_* | \mathcal{D})} [H(p(\mathbf{o}_C^k | \mathcal{D}, \mathbf{x}_*))]$ ). To evaluate (7), we approximate the integration over the maximizer  $\mathbf{x}_* \in \mathcal{X}$  as a summation over a finite set  $\mathcal{X}_* \subset \mathcal{X}$  of possible maximizers. If  $\mathcal{X}$  is discrete and  $|\mathcal{X}|$  is sufficiently small, we can set  $\mathcal{X}_* = \mathcal{X}$ . On the other hand, we can construct  $\mathcal{X}_*$  by optimizing function samples drawn from the GP posterior belief given  $\mathcal{D}$  (Hernández-Lobato, Hoffman, and Ghahramani 2014; Wang and Jegelka 2017; Rahimi and Recht 2008). For high dimensional problems, additive GP can be employed, as explained in (Wang and Jegelka 2017). Given the finite set  $\mathcal{X}_*$  of possible maximizers, (7) can be expressed as follows:

$$\begin{aligned} I(\mathbf{o}_C^k; \mathbf{x}_* | \mathcal{D}) &\approx \sum_{\mathbf{x}_* \in \mathcal{X}_*} p(\mathbf{x}_* | \mathcal{D}) \sum_{\mathbf{o}_C^k} p(\mathbf{o}_C^k | \mathcal{D}, \mathbf{x}_*) \log \frac{p(\mathbf{o}_C^k | \mathcal{D}, \mathbf{x}_*)}{p(\mathbf{o}_C^k | \mathcal{D})} \end{aligned} \quad (8)$$

where the probabilities are estimated with sampling in the following procedure:

1. Draw  $n$  samples  $\mathbf{f}_{\mathcal{C} \cup \mathcal{X}_*} \sim p(\mathbf{f}_{\mathcal{C} \cup \mathcal{X}_*} | \mathcal{D})$  of function values at  $\mathcal{X}_*$  and  $\mathcal{C}$  where  $p(\mathbf{f}_{\mathcal{C} \cup \mathcal{X}_*} | \mathcal{D})$  is the density of the multivariate Gaussian distribution specified in (6).
2. Estimate the posterior probability of  $\mathbf{x}_*$  given  $\mathcal{D}$  as  $p(\mathbf{x}_* | \mathcal{D}) = n^{-1} \sum_{\mathbf{f}_{\mathcal{X}_*}} \mathbb{1}_{\mathbf{x}_* = \arg\max \mathbf{f}_{\mathcal{X}_*}}$  where the summation is taken over function samples at  $\mathcal{X}_*$  in step 1.
3. Estimate the joint posterior probability of  $\mathbf{o}_C^k$  and  $\mathbf{x}_*$  given  $\mathcal{D}$  as  $p(\mathbf{o}_C^k, \mathbf{x}_* | \mathcal{D}) = n^{-1} \sum_{\mathbf{f}_{\mathcal{C} \cup \mathcal{X}_*}} \mathbb{1}_{\mathbf{x}_* = \arg\max \mathbf{f}_{\mathcal{X}_*}} p(\mathbf{o}_C^k | \mathbf{f}_{\mathcal{C}})$  where the summation is taken over function samples at  $\mathcal{C} \cup \mathcal{X}_*$  obtained in step 1 and the likelihood  $p(\mathbf{o}_C^k | \mathbf{f}_{\mathcal{C}})$  is described in Sec. 2.

4. Estimate the posterior probability of the observation  $\mathbf{o}_C^k$  given  $\mathcal{D}$  as  $p(\mathbf{o}_C^k|\mathcal{D}) = \sum_{\mathbf{x}_* \in \mathcal{X}_*} p(\mathbf{o}_C^k, \mathbf{x}_*|\mathcal{D})$  where  $p(\mathbf{o}_C^k, \mathbf{x}_*|\mathcal{D})$  is obtained in step 3.
5. Estimate the posterior probability of the observation  $\mathbf{o}_C^k$  given  $\mathcal{D}$  and the maximizer  $\mathbf{x}_*$  as  $p(\mathbf{o}_C^k|\mathcal{D}, \mathbf{x}_*) = p(\mathbf{o}_C^k, \mathbf{x}_*|\mathcal{D})/p(\mathbf{x}_*|\mathcal{D})$  where  $p(\mathbf{o}_C^k, \mathbf{x}_*|\mathcal{D})$  and  $p(\mathbf{x}_*|\mathcal{D})$  are obtained in steps 3 and 2, respectively.

Note that the number of possible top-1 rankings  $\mathbf{o}_C^1$  only grows linearly w.r.t.  $|\mathcal{C}|$ . So, the evaluation of MPES can scale well to large  $|\mathcal{C}|$  for  $k = 1$ . When  $k > 1$ , the number of possible  $\mathbf{o}_C^k$  grows exponentially w.r.t.  $|\mathcal{C}|$ . So, the cost of evaluating MPES is dominated by the sum over  $|\mathcal{C}|!/(|\mathcal{C}|-k)!$  possible observations. Therefore, we mainly focus on a small  $|\mathcal{C}|$  or  $k = 1$  such as  $|\mathcal{C}| = 4$  in our experiments where we enumerate all possible  $\mathbf{o}_C^k$  to compute MPES. In this case, we search for  $\operatorname{argmax}_{\mathcal{C} \in \mathcal{X}^{|\mathcal{C}|}} I(\mathbf{o}_C^k; \mathbf{x}_*|\mathcal{D})$  by randomly selecting a number of subsets of  $|\mathcal{C}|$  inputs in  $\mathcal{X}$ , which is empirically shown to significantly outperform EI and DTS in our experiments (Sec. 5). Alternatively, DIRECT (Jones, Perttunen, and Stuckman 1993) may be used to optimize MPES. Nonetheless, searching over a high dimensional space is a challenging problem. A potential direction to evaluate MPES for large  $|\mathcal{C}|$  and  $k > 1$  is to express (8) as

$$I(\mathbf{o}_C^k; \mathbf{x}_*|\mathcal{D}) \approx \sum_{\mathbf{x}_* \in \mathcal{X}_*} p(\mathbf{x}_*|\mathcal{D}) \mathbb{E}_{p(\mathbf{o}_C^k|\mathcal{D}, \mathbf{x}_*)} [\log p(\mathbf{o}_C^k|\mathcal{D}, \mathbf{x}_*)/p(\mathbf{o}_C^k|\mathcal{D})]$$

where the expectation  $\mathbb{E}_{p(\mathbf{o}_C^k|\mathcal{D}, \mathbf{x}_*)}$  is approximated by stochastic sampling, i.e., drawing a number of  $\mathbf{o}_C^k$  following  $p(\mathbf{o}_C^k|\mathcal{D}, \mathbf{x}_*)$ . It is performed by sampling  $\mathbf{o}_C^k$  from  $p(\mathbf{o}_C^k|\mathcal{D}, \mathbf{x}_*, \mathbf{f}_{\mathcal{C} \cup \mathcal{X}_*}) = p(\mathbf{o}_C^k|\mathbf{f}_{\mathcal{C} \cup \mathcal{X}_*})$  (i.e., the likelihood function in Sec. 2) where  $\mathbf{f}_{\mathcal{C} \cup \mathcal{X}_*}$  are samples in the above step 1 s.t.  $\mathbf{x}_* = \operatorname{argmax}_{\mathcal{X}_*} \mathbf{f}_{\mathcal{X}_*}$ . The rationale is to estimate the expectation with samples  $\mathbf{o}_C^k$  of high probabilities  $p(\mathbf{o}_C^k|\mathcal{D}, \mathbf{x}_*)$  instead of with all of the possible  $\mathbf{o}_C^k$  for large  $|\mathcal{C}|$  and  $k > 1$ .

While joint optimization over the query requires searching over a large space (i.e.,  $\mathcal{X}^{|\mathcal{C}|}$ ) as compared to optimizing the inputs of a query one a time greedily (i.e.,  $\mathcal{X}$ ) in (Brochu, Cora, and de Freitas 2010; Dewancker, Bauer, and McCourt 2017; González et al. 2017), the latter ignores the correlation between function values evaluated at different inputs of the query when selecting the first input. This leads to inferior performance, as shown empirically in our experiments (Sec. 5). Since the primary goal of BO is to reduce the cost of obtaining the query observation, the BO performance should not be sacrificed for the cost of optimizing the acquisition function. This is the motivation for us to develop a BO algorithm that is capable of jointly optimizing over the query to improve the quality of the queries.

**Example 1 (Exploitation vs. exploration)** Fig. 2c shows the MPES values for pairs of inputs in the domain  $\mathcal{X}$  given the GP posterior belief in Fig. 2a. The MPES values are symmetric about the line  $x_0 = x_1$ , which is expected as the role of  $x_0$  and  $x_1$  are interchangeable. There are two regions with high MPES values, which are annotated as exploitation and exploration. The exploitation region is the region where the

posterior means of both inputs in the pair are large. These inputs also have high probabilities of being the maximizer in Fig. 2b. The exploration region is the region where the posterior mean of one input in the pair is large (i.e., its probability of being the maximizer is high in Fig. 2b) and the other input is far away from all inputs in the observed pairwise preferences (i.e., its probability of being the maximizer is not high in Fig. 2b). Thus, MPES is able to balance exploration and exploitation naturally without any explicit modification.

The joint optimization over the query and the exploration-exploitation trade-off of MPES contrast with existing approaches: *expected improvement* (EI) (Brochu, Cora, and de Freitas 2010; Dewancker, Bauer, and McCourt 2017) and *dueling-Thompson sampling* (DTS) (González et al. 2017). In EI, the two inputs of the query (i.e., a pair of inputs) are selected independently: an input maximizing EI and the other input maximizing  $\mu_{\mathbf{x}}$ . Note that EI potentially leads to excessive exploitation (González et al. 2017). On the other hand, DTS selects inputs of a query one at a time: the first input maximizing the soft-Copeland score and the second input maximizing the variance of the preference given the first input. Furthermore, exploration is explicitly introduced by selecting the first input based on only 1 sample from the GP belief using continuous Thompson sampling.

## 5 Experiments and Discussion

In this section, we empirically demonstrate (a) the performance of our MPES in comparison with existing methods: *expected improvement* (EI) (Mockus, Tiešis, and Žilinskas 1978) and *dueling-Thompson sampling* (DTS) (González et al. 2017) using pairwise preferences in Sec. 5.1, (b) the performance of MPES and the model of ties in Sec. 5.2, and (c) the performance of MPES with top- $k$  ranking in Sec. 5.3. For the experimental results of MPES, we label MPES with the values of  $k$ ,  $|\mathcal{C}|$ , and  $\delta$ : For example, MPES with  $k = 1$ ,  $|\mathcal{C}| = 3$ , and  $\delta = 0$  (i.e., top-1 ranking of a set of 3 inputs without ties) is labeled as *top-1 of 3  $\delta = 0$* . When  $\delta > 0$  (i.e., ties exist),  $\delta$  is unknown to our model and is obtained by optimizing (5). To evaluate MPES, we set  $|\mathcal{X}_*|$  to 20 and the number of samples is  $n = 1000$ . The code is available at <https://github.com/sebtsh/Top-k-Ranking-Bayesian-Optimization>.

Following the work of Hernández-Lobato, Hoffman, and Ghahramani (2014), we compute the *immediate regret* in each BO iteration as the performance metric. For synthetic functions, it is the difference between the global maximum value  $\max_{\mathbf{x} \in \mathcal{X}} f_{\mathbf{x}}$  of the objective function and the function value at an estimate of the maximizer from the GP belief. This estimate of the maximizer is the maximizer of the GP posterior mean function (i.e.,  $\operatorname{argmax}_{\mathbf{x} \in \mathcal{X}} \mu_{\mathbf{x}}$ ) for MPES and EI, while it is the maximizer of the soft-Copeland score for DTS. A smaller immediate regret is preferred and indicates higher query efficiency. We repeat each experiment 10 times to plot the average and standard error of the immediate regret.

These acquisition functions are evaluated on 3 synthetic benchmark functions with varying levels of difficulty and number of dimensions: (a) the simple 1-D Forrester function (Forrester, Sobester, and Keane 2008), (b) the 2-D *six-hump*

*camel* (SHC) function (6 local minima) with the input domain restricted to  $[-1.5, 1.5]$  in each dimension, and (c) the 3-D Hartmann function.<sup>3</sup> These functions are originally modeled for finding the global minimum. However, since throughout this work, we have framed the problem as one of finding the global maximum, we take the negative values of these functions instead. The numbers of initial observations provided to the BO algorithms are 5, 6, and 12 for experiments with the Forrester, SHC, and Hartmann functions, respectively. We also perform experiments on the following 2 real-world datasets:

**CIFAR-10 dataset** The input domain consists of 50000 training images of the CIFAR-10 dataset (Krizhevsky 2009) which includes  $32 \times 32$  colour images in 10 classes. We use the following ground truth ranking of preference between classes: 7 (horse)  $\prec$  6 (frog)  $\prec$  5 (dog)  $\prec$  4 (deer)  $\prec$  3 (cat)  $\prec$  2 (bird)  $\prec$  9 (truck)  $\prec$  8 (ship)  $\prec$  1 (automobile)  $\prec$  0 (airplane). The objective is to identify the most preferred class through observing preferences/rankings between different images. Given the GP posterior belief, the most preferred class is defined as the class where the average of the posterior mean of all images in the class is the maximum. The immediate regret is defined as the distance from the most preferred class given the GP posterior belief to class 0 (i.e., airplane) in the ground truth ranking. For example, the immediate regret is 3 if the most preferred class given the GP posterior belief is class 9 (i.e., truck). So, the immediate regret is an integer in the range  $[0, 9]$ . We reduce the dimensionality of CIFAR-10 dataset to an embedding space of 2 dimensions with a combination of transfer learning from a CNN and a UMAP reduction (McInnes et al. 2018). The embedding is visualized by plotting a smooth function of the ground truth ranking with the embeddings of images as the inputs in Fig. 3. We observe that this embedding separates the 10 classes of CIFAR-10 into different clusters while preserving the relative distance between images such that visually similar images are close in the embedding, and vice versa. For example, classes 3 (i.e., cat) and 5 (i.e., dog) are close to each other due to cats and dogs being visually similar. The objective function maps the 2-D embedding of an image to the order of its class in the ground truth ranking subtracted by 5 (e.g., the function value of a horse image is  $1 - 5 = -4$ ). Six initial observations are provided to the BO algorithms.

**SUSHI preference dataset** Inspired from our dining example in Sec. 1, this experiment is about learning the most preferred type of sushi through ranking observations. The objective function is generated from the real-world SUSHI preference dataset (Kamishima 2003), which is widely used for the evaluation of preference and ranking methods (Khetan and Oh 2016; Vitelli et al. 2018). It consists of data for 100 kinds of sushi and 5000 user ratings of subsets of sushi. The input  $x$  consists of 6 features of the sushi. The objective function is obtained by scaling and shifting the average rating scores of users (which represents their average opinion) to the range  $[-4, 5]$  that is similar to the CIFAR-10 experiment.

<sup>3</sup>Available at [http://www-optima.amp.i.kyoto-u.ac.jp/member/student/hedar/Hedar\\_files/TestGO\\_files/Page1488.htm](http://www-optima.amp.i.kyoto-u.ac.jp/member/student/hedar/Hedar_files/TestGO_files/Page1488.htm).

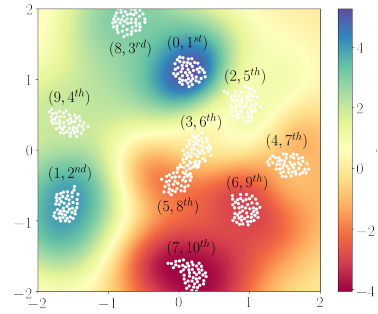


Figure 3: Plot of a smooth function of the ground truth ranking with the 2-D embeddings of a subset of the CIFAR-10 dataset. Tuples in each cluster indicate the class number, followed by the order of the class in the ground truth ranking.

The immediate regret is calculated as the distance between the BO algorithm’s best guess of the most preferred sushi and the actual most preferred sushi in the ground truth ranking (based on the average rating scores). For example, if salmon sushi has rank 10 in the ground truth ranking and the BO algorithm’s best guess of the most preferred sushi at a timestep is salmon sushi, the algorithm’s immediate regret at that timestep would be 9 because salmon sushi is 9 places away from the top. In these experiments, there are 10 initial observations provided to the BO algorithms.

### 5.1 BO with Pairwise Preferences

Since the existing EI and DTS approaches are only able to handle pairwise preferences, we compare the performance of our MPES with EI and DTS using pairwise preferences here. The immediate regrets for the Forrester, SHC, Hartmann, CIFAR-10, and SUSHI are shown in Fig. 4. It can be observed that MPES consistently outperforms both EI and DTS in these experiments. DTS outperforms EI in optimizing the Forrester function with 1-D inputs (Fig. 4a). However, as the input dimension increases to 3 in the optimization of the Hartmann function (Fig. 4c) and to 6 in the optimization problem with the SUSHI dataset (Fig. 4e), DTS is outperformed by EI. This is due to the disadvantage of the model used by DTS whose input dimension is doubled with respect to the original dimension of the problem, as discussed in Sec. 3.

### 5.2 BO with Ties

In this subsection, we empirically show the competitive performance of MPES with tie observations by comparing with the existing DTS and EI using pairwise preferences. On top of that, we let DTS and EI have an unfair advantage over MPES by having access to strict preferences (i.e., no tie) regardless of how close the utility values of the input pair are, while MPES can only receive a tie observation (i.e., there is no information about the preferred input in the pair) if the difference in the utility values of the input pair is less than  $\delta$ . Even so, MPES with the model of ties is able to outperform DTS and EI both in optimizing synthetic benchmark functions and on the real-world datasets in Figs. 5a-e. This empirically shows the performance of both the model of ties in the likelihood (Sec. 2.2) and the performance of MPES.

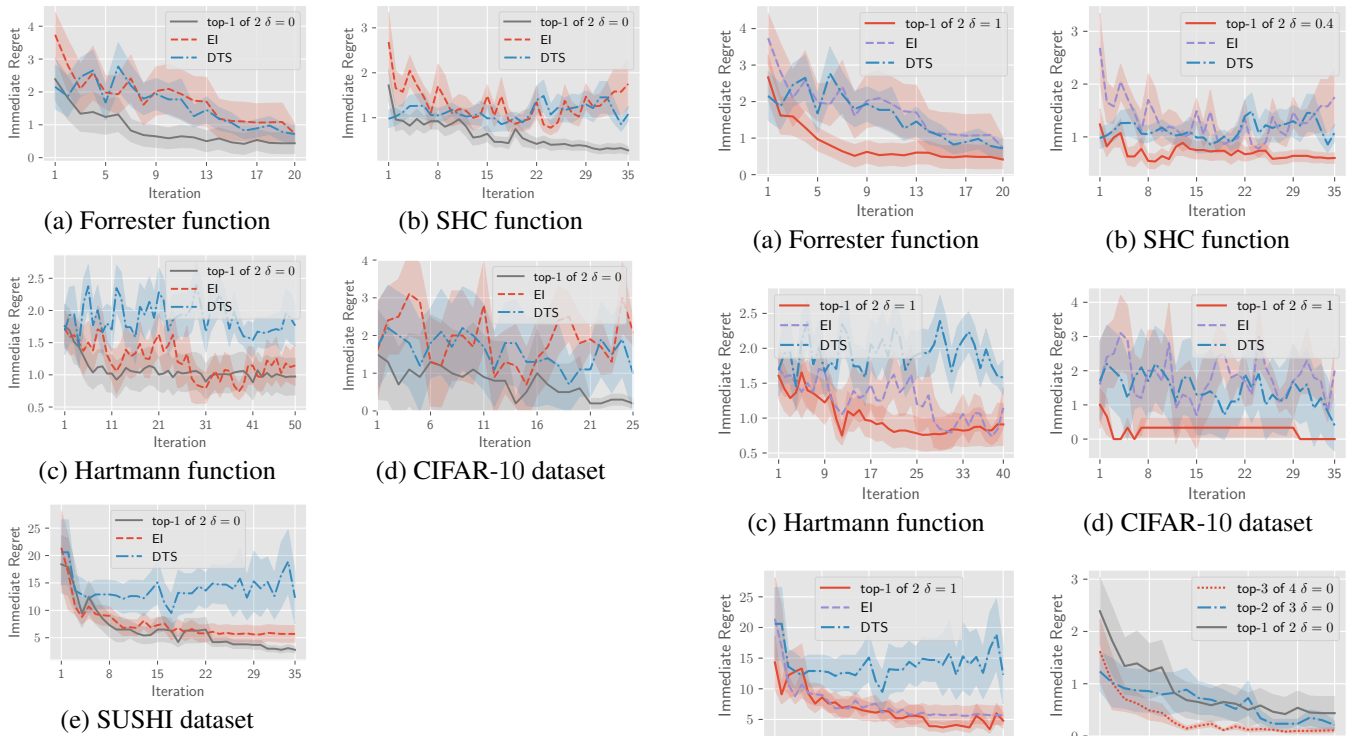


Figure 4: Plots of immediate regrets for experiments with pairwise preferences.

### 5.3 BO with Rankings

In this subsection, we empirically illustrate the advantage of ranking observations over pairwise preferences. In particular, as the rankings give more information about the GP posterior belief of the objective function, BO with ranking observation is expected to outperform BO with pairwise preferential observation given the same number of queries. This is illustrated in Figs. 5f-j where BO with  $|\mathcal{C}| > 2$  outperforms BO with  $|\mathcal{C}| = 2$  (i.e., pairwise preferences). Furthermore, as  $|\mathcal{C}|$  increases, the performance of our algorithm improves.

## 6 Conclusion

This paper describes a principled approach to top- $k$  BO in both modeling the posterior belief of the objective function and formulating the acquisition function. Inspired by the classic multinomial logit model, our model is capable of handling real-world observations including top- $k$  rankings of inputs and the existence of ties. Furthermore, based on an information-theoretic measure, we design the MPES acquisition function that is capable of guiding the query selection through jointly optimizing all inputs of a query and balancing the exploration-exploitation trade-off. Our new model and MPES are empirically demonstrated to have superior performance compared with existing approaches in several synthetic benchmark functions, CIFAR-10 dataset, and SUSHI preference dataset. For future work, we plan to generalize MPES to nonmyopic BO (Kharkovskii, Ling, and Low 2020; Ling, Low, and Jaillet 2016), batch BO (Daxberger and Low 2017), high-dimensional BO (Hoang, Hoang, and Low 2018),

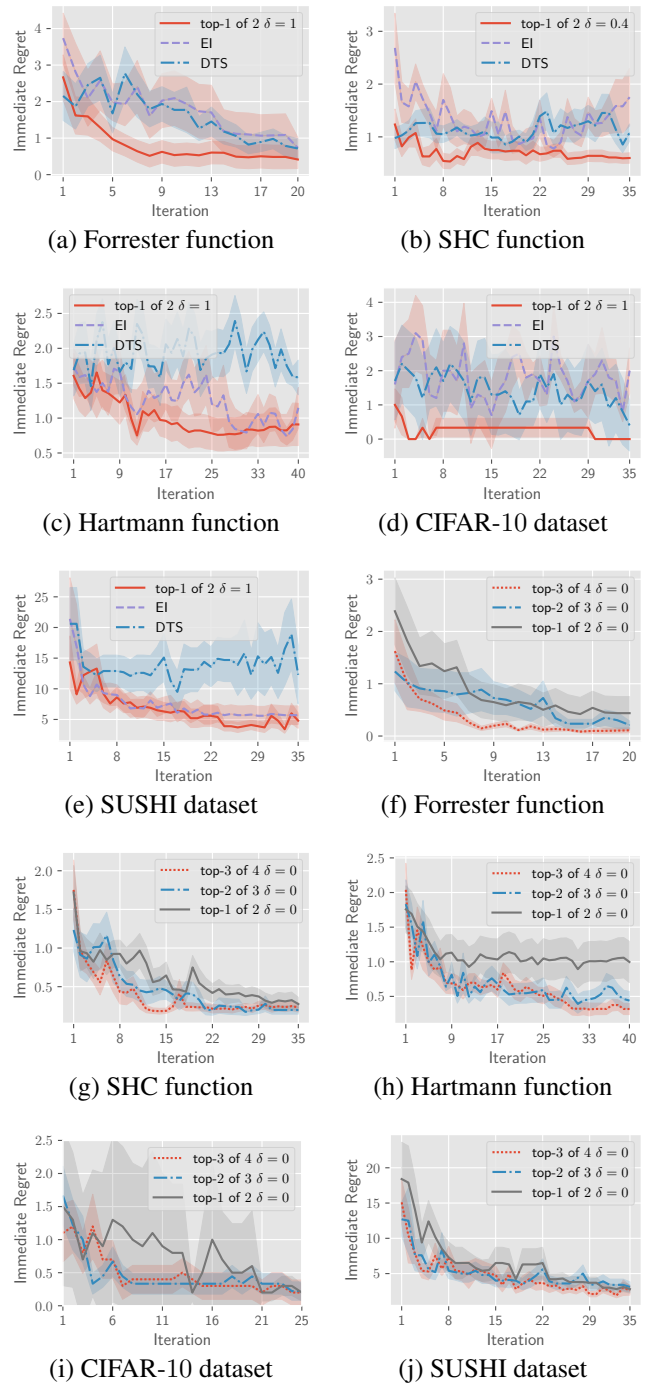


Figure 5: Plots of immediate regrets for experiments with (a-e) pairwise preferences and (f-j) rankings. The observation of MPES can include ties while that of EI and DTS only include strict preferences (i.e., no tie). The threshold  $\delta$  is learned from observation.

private outsourced BO (Kharkovskii, Dai, and Low 2020), and multi-fidelity BO (Zhang et al. 2017; Zhang, Dai, and Low 2019) settings and incorporating early stopping (Dai et al. 2019) and recursive reasoning (Dai et al. 2020).

## Ethical Impact

There are many real-world applications (e.g., food/movie/music preference, art aesthetics, interior design, and place of interests (tourism)) where the objective function cannot be directly evaluated. In these applications, top- $k$  ranking BO is a promising optimization method (e.g., finding the best food recipe, the most pleasing design, and the most attractive place for tourists) with a limited budget of queries/trials. Furthermore, the possibility of different types of observation in our models provides more options to design the data collection process, which potentially improves the user experience.

There are also several considerations in applying our model. The negative impacts of our method can happen due to the poor performance when the underlying assumptions of our model (e.g., the Gumbel noise and the independence of irrelevant alternatives (inherent in the logit model)) are violated. Therefore, an application is required to examine if these assumptions are satisfied. In case they do not hold, further investigation into alternative models is necessary. For applications with constraints (e.g., a combination of ingredients can cause food poisoning or be unsafe during pregnancy), we need to extend the current work to incorporate the constraints into the optimization process.

In brief, we believe our work is beneficial to the society when the above issues are taken into consideration. It is a step forward to extend the use of BO to daily applications.

## Acknowledgments

This research/project is supported by A\*STAR under its RIE2020 Advanced Manufacturing and Engineering (AME) Industry Alignment Fund – Pre Positioning (IAF-PP) (Award A19E4a0101).

## References

- Brochu, E.; Cora, V. M.; and de Freitas, N. 2010. A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. arXiv:1012.2599.
- Busa-Fekete, R.; Hüllermeier, E.; and Mesaoudi-Paul, A. E. 2018. Preference-based online learning with dueling bandits: A survey. arXiv:1807.11398.
- Cantillo, V.; Amaya, J.; and Ortúzar, J. D. 2010. Thresholds and indifference in stated choice surveys. *Transport. Res. B-Meth.* 44(6): 753–763.
- Chu, W.; and Ghahramani, Z. 2005. Preference learning with Gaussian processes. In *ICML*, 137–144.
- Dai, Z.; Chen, Y.; Low, B. K. H.; Jaillet, P.; and Ho, T.-H. 2020. R2-B2: Recursive reasoning-based Bayesian optimization for no-regret learning in games. In *Proc. ICML*, 2291–2301.
- Dai, Z.; Yu, H.; Low, B. K. H.; and Jaillet, P. 2019. Bayesian optimization meets Bayesian optimal stopping. In *Proc. ICML*, 1496–1506.
- Daxberger, E. A.; and Low, B. K. H. 2017. Distributed Batch Gaussian process optimization. In *Proc. ICML*, 951–960.
- Dewancker, I.; Bauer, J.; and McCourt, M. 2017. Sequential preference-based optimization. In *Proc. NeurIPS Workshop on Bayesian Deep Learning*.
- Forrester, A. I. J.; Sobester, A.; and Keane, A. J. 2008. *Engineering Design via Surrogate Modelling: A Practical Guide*. John Wiley & Sons.
- González, J.; Dai, Z.; Damianou, A.; and Lawrence, N. D. 2017. Preferential Bayesian optimization. In *Proc. ICML*, 1282–1291.
- Hennig, P.; and Schuler, C. J. 2012. Entropy search for information-efficient global optimization. *JMLR* 13: 1809–1837.
- Hernández-Lobato, J. M.; Hoffman, M. W.; and Ghahramani, Z. 2014. Predictive entropy search for efficient global optimization of black-box functions. In *Proc. NeurIPS*, 918–926.
- Hoang, T. N.; Hoang, Q. M.; and Low, B. K. H. 2018. Decentralized high-dimensional Bayesian optimization with factor graphs. In *Proc. AAAI*, 3231–3238.
- Jones, D. R.; Perttunen, C. D.; and Stuckman, B. E. 1993. Lipschitzian optimization without the Lipschitz constant. *J. Optimization Theory and Applications* 79(1): 157–181.
- Kamishima, T. 2003. Nantonac collaborative filtering: Recommendation based on order responses. In *Proc. ACM SIGKDD*, 583–588.
- Kharkovskii, D.; Dai, Z.; and Low, B. K. H. 2020. Private outsourced Bayesian optimization. In *Proc. ICML*, 5231–5242.
- Kharkovskii, D.; Ling, C. K.; and Low, B. K. H. 2020. Nonmyopic Gaussian process optimization with macro-actions. In *Proc. AISTATS*, 4593–4604.
- Khetan, A.; and Oh, S. 2016. Data-driven rank breaking for efficient rank aggregation. *JMLR* 17(193): 1–54.
- Krizhevsky, A. 2009. *Learning multiple layers of features from tiny images*. Master’s thesis, Department of Computer Science, University of Toronto.
- Ling, C. K.; Low, B. K. H.; and Jaillet, P. 2016. Gaussian process planning with Lipschitz continuous reward functions: Towards unifying Bayesian optimization, active learning, and beyond. In *Proc. AAAI*, 1860–1866.
- Luce, R. D. 1959. *Individual Choice Behavior: A Theoretical Analysis*. John Wiley & Sons.
- Marschak, J. 1959. Binary choice constraints on random utility indicators. Cowles foundation discussion paper no. 74, Cowles Foundation for Research in Economics, Yale University.
- McFadden, D. 1974. The measurement of urban travel demand. *J. Public Economics* 3(4): 303–328.
- McInnes, L.; Healy, J.; Saul, N.; and Großberger, L. 2018. UMAP: Uniform manifold approximation and projection. *The Journal of Open Source Software* 3(29): 861.
- Mockus, J.; Tiešis, V.; and Žilinskas, A. 1978. The application of Bayesian methods for seeking the extremum. In Dixon, L. C. W.; and Szegő, G. P., eds., *Towards Global Optimization* 2, 117–128. North-Holland Publishing Company.

- Nguyen, Q. P.; Tay, S.; Low, B. K. H.; and Jaillet, P. 2020. Top- $k$  ranking Bayesian optimization. *arXiv:2012.10688*.
- Plackett, R. L. 1975. The analysis of permutations. *J. R. Statist. Soc. C* 24(2): 193–202.
- Quiñonero-Candela, J.; and Rasmussen, C. E. 2005. A unifying view of sparse approximate Gaussian process regression. *JMLR* 6: 1939–1959.
- Rahimi, A.; and Recht, B. 2008. Random features for large-scale kernel machines. In *Proc. NeurIPS*, 1177–1184.
- Rasmussen, C. E.; and Williams, C. K. I. 2006. *Gaussian Processes for Machine Learning*. MIT Press.
- Ru, B.; McLeod, M.; Granzio, D.; and Osborne, M. A. 2018. Fast information-theoretic Bayesian optimisation. In *Proc. ICML*, 4381–4389.
- Shah, A.; and Ghahramani, Z. 2015. Parallel predictive entropy search for batch global optimization of expensive objective functions. In *Proc. NeurIPS*, 3330–3338.
- Sui, Y.; Zhuang, V.; Burdick, J. W.; and Yue, Y. 2017. Multi-dueling bandits with dependent arms. In *Proc. UAI*.
- Train, K. E. 2009. *Discrete Choice Methods with Simulation*. Cambridge Univ. Press.
- Villemonteix, J.; Vazquez, E.; and Walter, E. 2009. An informational approach to the global optimization of expensive-to-evaluate functions. *J. Glob. Optim.* 44(4): 509–534.
- Vitelli, V.; Sørensen, Ø.; Crispino, M.; Frigessi, A.; and Arjas, E. 2018. Probabilistic preference learning with the Mallows rank model. *JMLR* 18(158): 1–49.
- Wang, Z.; and Jegelka, S. 2017. Max-value entropy search for efficient Bayesian optimization. In *Proc. ICML*, 3627–3635.
- Zhang, Y.; Dai, Z.; and Low, B. K. H. 2019. Bayesian optimization with binary auxiliary information. In *Proc. UAI*.
- Zhang, Y.; Hoang, T. N.; Low, B. K. H.; and Kankanhalli, M. 2017. Information-based multi-fidelity Bayesian optimization. In *Proc. NIPS Workshop on Bayesian Optimization*.