Bayesian Optimisation with Pairwise Preferential Returns

Javier González
Amazon.com, Cambridge, UK
gojav@amazon.com

Zhenwen Dai
Amazon.com, Cambridge, UK
zhenwend@amazon.com

Andreas Damianou
Amazon.com, Cambridge, UK
damianou@amazon.com

Neil D. Lawrence
Amazon.com, Cambridge, UK
University of Sheffield
lawrennd@amazon.com

Abstract

Bayesian optimisation has emerged during the last few years as an effective approach to optimise *black-box* functions where direct queries of the objective are expensive. In many real applications, however, pairwise preferences rather than direct feedback values are available. Such scenarios arise, for instance, in A/B tests or recommendation systems. We present BOPPER, *Bayesian Optimisation with Pairwise PrEferential Returns*, a new global optimisation approach able to find the optimum of a latent function that can only be queried through pairwise comparisons, the so-called *duels*. BOPPER generalises previous discrete duelling approaches by modelling the probability of the the winner of each duel by means of Gaussian process model with a Bernoulli likelihood. The latent preference function is used to define the Copeland Expected Improvement (CEI), a new acquisition function tailored to this scenario. We illustrate the benefits of BOPPER in a variety of experiments.[1]

1 Introduction

Let $g : \mathcal{X} \to \Re$ be well-behaved *black-box* function defined on a bounded subset $\mathcal{X} \subseteq \Re^q$. We are interested in solving the global optimisation problem of finding

$$\mathbf{x}_{min} = \arg\min_{\mathbf{x}\in\mathcal{X}} g(\mathbf{x}). \quad (1)$$

We assume that g is not directly accessible and that queries to g can only be done in pairs of points or *duels* $[\mathbf{x}, \mathbf{x}'] \in \mathcal{X} \times \mathcal{X}$ from which binary feedback $\{0, 1\}$ that represents whether or not $\mathbf{x}$ is preferred over $\mathbf{x}'$ (has lower value) is obtained[2]. In the sequel we will consider that $\mathbf{x}$ is the winner of the duel if the output is $\{1\}$ and that $\mathbf{x}'$ wins the duel otherwise if the output is $\{0\}$. The goal here is to find $\mathbf{x}_{min}$ by reducing as much as possible the number of performed duels.

Our setup is different to the one typically used in Bayesian optimisation where direct feedback from g in single locations of the domain is available [Jones, 2001, Snoek et al., 2012]. However, although the scenario described in this work has not received a wider attention, there exist a variety of real wold scenarios in which the objective function needs to be optimized via preferential returns. Most cases involve modeling *latent human preferences*, such as examples in the web design via A/B testing or the use of recommender systems.

[1] Work done while all the authors were at the University of Sheffield.

[2] In the sequel we use $[\mathbf{x}, \mathbf{x}']$ to represent the vector resulting of concatenating both elements involved in the duel.

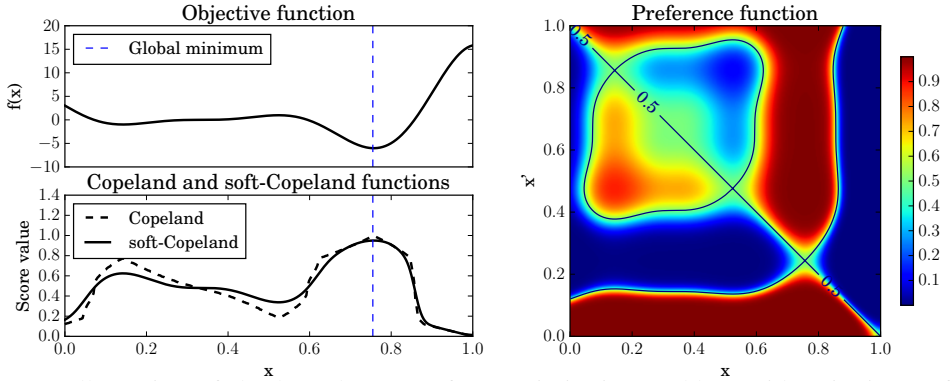

Figure 1: Illustration of the key elements of an optimisation problem with pairwise preferential returns in a one-dimensional example (Forrester function, see Section 4 for details). *Top-left*: objective function to minimise. This function is only accessible through pairwise comparisons of inputs $\mathbf{x}$ and $\mathbf{x}'$. *Right*: true preference function $\pi_f([\mathbf{x}, \mathbf{x}'])$. Note that, by symmetry, $\pi_f([\mathbf{x}, \mathbf{x}']) = 1 - \pi_f([\mathbf{x}', \mathbf{x}])$. *Bottom left*: The normalised Copeland's and soft-Copeland function whose maximum is located at the same point of the minimum of f.

Optimisation methods for pairwise preferences have been already studied in the armed-bandits context [Yuea et al., 2012]. Zoghi et al. [2014] propose a method for the K-armed duelling bandit problem based on the Upper Confidence Bound algorithm. Jamieson et al. [2015] study the problem by allowing noise comparisons between the duels. Zoghi et al. [2015] choose actions using contextual information. Dudík et al. [2015] study the Copeland's dueling bandits, a case in which a Condorcet winner, or an arm that uniformly wins the duels with all the other arms may not exist. Szörényi et al. [2015] study Online Rank Elicitation problem in the duelling bandits setting. An analysis on Thompson sampling in duelling bandits is done bu Wu et al. [2016]. Yue and Joachims [2011] proposes a method that does not need transitivity and comparison outcomes to have independent and time-stationary distributions. From a modelling perspective preferential learning has also been studied [Chu and Ghahramani, 2005].

2 Background and Approach

The approach followed in this work is inspired on the work of Ailon et al. [2014] in which cardinal bandits are reduced to ordinal ones. Similarly, here we focus on the idea of reducing the choice of the best duel to some optimisation problem defined on $\mathcal{X}$ whose solution is the same as (1).

We assume that each duel $[\mathbf{x}, \mathbf{x}']$ incurs in a joint loss $f([\mathbf{x}, \mathbf{x}'])$ that is never directly observed. Instead, the feedback after each pair is proposed is a binary return $y \in \{0, 1\}$ of which of the two locations is preferred. In this work we assume that $f([\mathbf{x}, \mathbf{x}']) = g(\mathbf{x}') - g(\mathbf{x})$, but other alternatives are possible. Note that the more $\mathbf{x}$ is preferred over $\mathbf{x}'$ the smaller is the loss.

The model of choice is a Bernoulli probability function $p(y = 1|[\mathbf{x}, \mathbf{x}']) = \pi_f([\mathbf{x}, \mathbf{x}'])$ and $p(y = 0|[\mathbf{x}, \mathbf{x}']) = \pi_f([\mathbf{x}', \mathbf{x}])$ where $\pi : \Re \times \Re \to [0, 1]$ is a link function. Via the latent loss, f maps each query $[\mathbf{x}, \mathbf{x}']$ to the probability of having a preference on the left input $\mathbf{x}$ over the right input $\mathbf{x}'$. The link function has the property that $\pi_f([\mathbf{x}, \mathbf{x}']) = 1 - \pi_f([\mathbf{x}', \mathbf{x}])$. A natural choice for π_f is the logistic function

$$\pi_f([\mathbf{x}', \mathbf{x}]) = \sigma(f([\mathbf{x}', \mathbf{x}])) = \frac{1}{1 + e^{-f([\mathbf{x}', \mathbf{x}])}}. \tag{2}$$

but others are possible. Note that for any duel $[\mathbf{x}, \mathbf{x}']$ in which $g(\mathbf{x}) \leq g(\mathbf{x}')$ it holds that $\pi_f([\mathbf{x}, \mathbf{x}']) \geq 0.5$. π_f is therefore a *preference function* that fully specified the problem.

Following the literature of raking methods, we introduce here the concept of *normalised Copeland score* as $S(\mathbf{x}) = \text{Vol}(\mathcal{X})^{-1} \int_{\mathcal{X}} \mathbb{I}_{\{\pi_f([\mathbf{x}, \mathbf{x}']) \geq 0.5\}} d\mathbf{x}'$ where $\text{Vol}(\mathcal{X}) = \int_{\mathcal{X}} d\mathbf{x}'$ is a normalizing constant that bounds $S(\mathbf{x})$ in the interval $[0, 1]$. If $\mathcal{X}$ is a finite set, the Copeland score is simply the proportion of duels that certain element $\mathbf{x}$ will win with probability larger than 0.5. Instead of the Copeland's score in this work we use a soft version of it, in which the probability function π_f is

Algorithm 1 The BOPPER algorithm.

Input: Dataset $\mathcal{D}_0 = \{[\mathbf{x}_i, \mathbf{x}'_i], y_i\}_{i=1}^N$ and number of remaining evaluations n.
for $j = 0$ **to** n **do**
1. Fit a GP with kernel k to $\mathcal{D}_j$ and learn the probability preferences function $\pi_{f,j}(\mathbf{x})$.
2. Obtain the Copeland function $C_j(\mathbf{x})$ by Montecarlo integration and compute $\mathbf{x}_j^\star$.
3. Select next duel $[\mathbf{x}_{j+1}, \mathbf{x}'_{j+1}]$ by maximizing CEI defined in (4).
4. Run the duel $[\mathbf{x}_{j+1}, \mathbf{x}'_{j+1}]$ and obtain y_{j+1}.
5. Augment the dataset $\mathcal{D}_{j+1} = \{\mathcal{D}_j \cup ([\mathbf{x}_{j+1}, \mathbf{x}'_{j+1}], y_{j+1})\}$.

end for
Fit a GP with kernel k to $\mathcal{D}_n$.
Returns: Report the current Condorcet's winner $\mathbf{x}_n^\star$.

integrated over $\mathcal{X}$ without further truncation. Formally, we define the soft-Copeland score as

$$C(\mathbf{x}) = \text{Vol}(\mathcal{X})^{-1} \int_{\mathcal{X}} \pi_f([\mathbf{x}, \mathbf{x}'])d\mathbf{x}' \tag{3}$$

which aims to capture the 'averaged' probability of being $\mathbf{x}$ the winner of a duel.

Following the armed-bandits literature, we say that $\mathbf{x}^\star$ is a *Condorcet winner* if it is point with maximal soft-Copeland score. It is possible to prove that if $\mathbf{x}^\star$ is a Condorcet winner with respect to the soft-Copeland score then it is a global minimum of f in $\mathcal{X}$. This implies that if by observing the results of a set of duels we can learn the preference function $\pi_f([\mathbf{x}', \mathbf{x}])$ then the optimisation problem of finding the minimum of f can be addressed by finding the Condorcet winner according to the Copeland score. See Figure 1 for an illustration of this property.

3 Considering preferential returns in Bayesian optimisation

3.1 Learning the preference function $\pi_f([\mathbf{x}, \mathbf{x}'])$ with Gaussian processes

Assume that N duels have been performed so far resulting in a dataset $\mathcal{D}_0 = \{[\mathbf{x}_i, \mathbf{x}'_i], y_i\}_{i=1}^N$ and that we can carry out n more before we have to report a solution to (1). We will denote by $\mathcal{D}_j$ the data set resulting of augmenting $\mathcal{D}_0$ with j new pairwise comparisons. Given $\mathcal{D}_j$, inference over the latent function f and its warped version π_f can be carried out by using Gaussian processes (GP) for classification [Rasmussen and Williams, 2005]. In a nutshell, a GP is a probably measure over functions such that any linear restriction is multivariate Gaussian. GPS are fully determined by a positive definite covariance operator and, in standard regression cases with Gaussian likelihoods, closed forms for the posterior mean and variance are available. In the classification context, the basic idea behind Gaussian process is to place a GP prior over some latent function f that captures the membership of the data to the two classes and to squash it through the logistic function to obtain some prior probability π_f. This is similar to (2) where now π_f is a random process, so it is f. Although in the regression context predictions are straightforward with a GP, in the classification context predictions are analytically intractable and either analytical approximations or Montecarlo sampling is needed. See [Rasmussen and Williams, 2005] for details.

3.2 Computing the soft-Copenland score and and the Condorcet winner

Denote by f_j the GP learnt once j duels have been performed and by $\pi_{f,j}(\mathbf{x})$ the corresponding squashed probability function. In this work we use Montecarlo integration to compute the Copeland score at the via $C_j(\mathbf{x}) = M^{-1} \sum_{k=1}^M \pi_{f,j}([\mathbf{x}, \mathbf{x}_k]))$ where $\mathbf{x}_1, \ldots, \mathbf{x}_M$ are a set of landmark points to perform the integration. The Condorcet winner (the point that is most likely the minimum in the j-th is computed by taking $\mathbf{x}_j^\star = \arg\max_{\mathbf{x} \in \mathcal{X}} C_j(\mathbf{x})$.

3.3 Copeland Expected Improvement

Denote by $c_j^\star = C_j(\mathbf{x}_j^\star)$, the 'value' of the Condorcet's winner at iteration j. For any new proposed duel $[\mathbf{x}, \mathbf{x}']$, two outcomes are possible. We denote by the $c_{j,\mathbf{x}}^\star$ the value of the estimated Condorcet

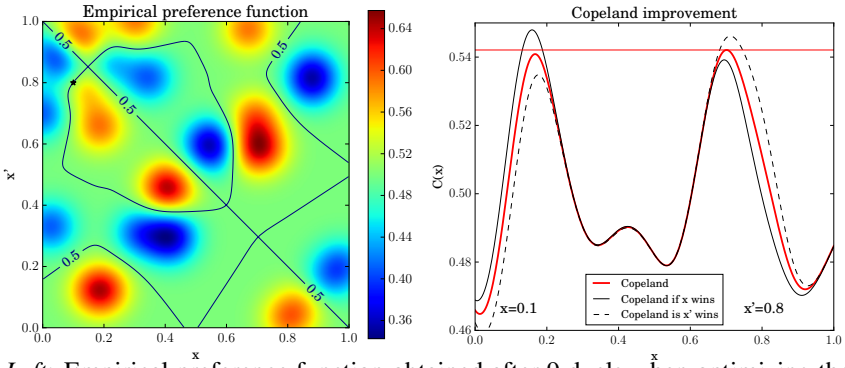

Figure 2: *Left*: Empirical preference function obtained after 9 duels when optimizing the Forrester function. The star at $[0.1, 0.8]$ represents a candidate point to perform a new duel. *Right*: Computed Copeland function and the two phantasized Copeland functions in terms on the result of the duel. Both potential outcomes improve the current best score. The value of the acquisition is a weighted average of the improvements.

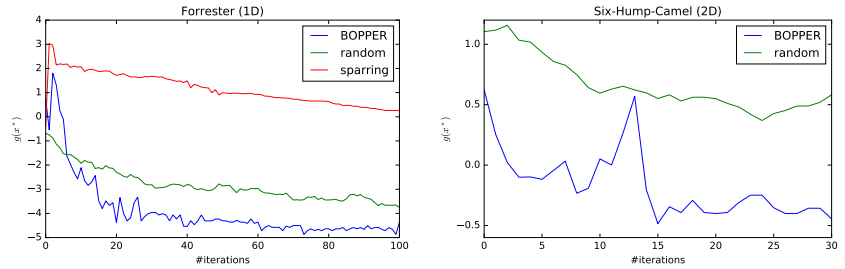

Figure 3: Averaged value of the objective in the current best value proposed by the BOPPER, the Sparring algorithm and a random duels search algorithms in two problems with preferential returns. BOPPER is shows the best performance in both cases. We omit the result of the Sparring algorithm in case of the Six-Hump camel function as the number of use iterations is smaller than the possible locations (and results are therefore uninterpretable) .

winner resulting of augmenting $\mathcal{D}_j$ with $\{[\mathbf{x}, \mathbf{x}'], 1\}$ and by $c^\star_{j,\mathbf{x}'}$ the equivalent value but augmenting the dataset with $\{[\mathbf{x}, \mathbf{x}'], 0\}$. We define the Copeland Expected Improvement at iteration j as:

$$CEI_j([\mathbf{x}, \mathbf{x}']) = \pi_{f,j}([\mathbf{x}, \mathbf{x}'])(c^\star_{j,\mathbf{x}} - c^\star_j)_+ + \pi_{f,j}([\mathbf{x}', \mathbf{x}])(c^\star_{j,\mathbf{x}'} - c^\star_j)_+ \tag{4}$$

where $(\cdot)_+ = \max(0, \cdot)$. The next duel is selected at the pair that maximizes the CEI. Intuitively, the CEI evaluated at $[\mathbf{x}, \mathbf{x}']$ is a weighted sum of the total increase of the best possible value of the Copeland score in the two possible outcomes of the duel. The weights are chosen to be the probability of the two outcomes, which is naturally estimated by $\pi_{f,j}$. See Figure 2 for details of how the acquisition is computed and Algorithm 1 for a systematic description of BOPPER.

4 Experiments

We compare BOPPER with the Sparring algorithm proposed in [Ailon et al., 2014] and a random policy (duels are selected using a uniform distribution). We test the methods in the optimisation of the Forrester (1D) and Six-Hump-Camel (2D) functions[3] using as model of choice for the output of the duels the framework described in Section 2. The search is performed in a uniform grid of 30 locations for the Forrester function and 64 for the Six-Hump camel. We assign to each problem a budget of 100 and 30 duels respectively after which the best location of the optimum should be reported. Each algorithm is run 10 times with different initial duels, which are kept the same across all methods. In Figure 3 we compare the methods by showing the averaged value of g at the proposed locations. BOPPER is best policy in all cases.

[3]https://www.sfu.ca/ ssurjano/optimisation.html

5 Conclusions

We have proposed a new method, BOPPER, for optimizing black-box functions in which only preferential returns are available. The new approach improve previous bandits alternatives by modeling the correlation between the candidates of to the optimum with a Gaussian process for classification. Future extensions of the work include improvements in the approximation of the Copeland function, further experimentation and a theoretical analysis of the convergence of the proposed method.

References

Nir Ailon, Zohar Shay Karnin, and Thorsten Joachims. Reducing dueling bandits to cardinal bandits. In *Proceedings of the 31th International Conference on Machine Learning, ICML 2014, Beijing, China, 21-26 June 2014*, pages 856–864, 2014.

Wei Chu and Zoubin Ghahramani. Preference learning with Gaussian processes. In *Proceedings of the 22Nd International Conference on Machine Learning*, ICML '05, pages 137–144, New York, NY, USA, 2005. ACM. ISBN 1-59593-180-5.

Miroslav Dudík, Katja Hofmann, Robert E. Schapire, Aleksandrs Slivkins, and Masrour Zoghi. Contextual dueling bandits. In *Proceedings of The 28th Conference on Learning Theory, COLT 2015, Paris, France, July 3-6, 2015*, pages 563–587, 2015.

Kevin G. Jamieson, Sumeet Katariya, Atul Deshpande, and Robert D. Nowak. Sparse dueling bandits. In *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics, AISTATS 2015, San Diego, California, USA, May 9-12, 2015*, 2015.

Donald R. Jones. A taxonomy of global optimization methods based on response surfaces. *Journal of global optimization*, 21(4):345383, 2001.

Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2005.

Jasper Snoek, Hugo Larochelle, and Ryan P. Adams. *Practical Bayesian optimization of machine learning algorithms*, page 29512959. 2012.

Balázs Szörényi, Róbert Busa-Fekete, Adil Paul, and Eyke Hüllermeier. Online rank elicitation for plackett-luce: A dueling bandits approach. In *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada*, pages 604–612, 2015.

Huasen Wu, Xin Liu, and R. Srikant. Double Thompson sampling for dueling bandits. *CoRR*, abs/1604.07101, 2016.

Yisong Yue and Thorsten Joachims. Beat the mean bandit. In *ICML*, pages 241–248, 2011.

Yisong Yuea, Josef Broderb, Robert Kleinbergc, and Thorsten Joachims. The k-armed dueling bandits problem. *Journal of Computer and System Sciences*, 78(5):1538 – 1556, 2012. ISSN 0022-0000. {JCSS} Special Issue: Cloud Computing 2011.

Masrour Zoghi, Shimon Whiteson, Remi Munos, and Maarten de Rijke. Relative upper confidence bound for the K-armed dueling bandit problem. In *ICML 2014: Proceedings of the Thirty-First International Conference on Machine Learning*, pages 10–18, June 2014.

Masrour Zoghi, Zohar S. Karnin, Shimon Whiteson, and Maarten de Rijke. Copeland dueling bandits. In *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada*, pages 307–315, 2015.