Submission and Formatting Instructions for International Conference on Machine Learning (ICML 2023)

Justin T. Chiu [1] **Wenting Zhao** [1] **Derek Chen** [2] **Alexander M. Rush** [1]

Abstract

Collaborative dialogue agents must balance communicative efficiency with task process for effective collaboration. We design a system that strategically asks questions while minimizing communication costs via symbolic planning. Our system consists of a reader, planner, and writer: The reader maps language to logical forms which inform the symbolic planner's next decision, represented as a logical form. The writer then converts the logical form to a response, taking into account any information not captured by the reader. Results on DialOp, negotiation in Deal-or-no-Deal.

1. Introduction

Our goal is to design robots that collaborate with humans to solve complex problems. Humans often know the problem specification, but may have difficulty solving the problem itself. Describing the problem in its entirety is expensive, requiring humans many words to convey. The marginal improvement from full information may not be worth the communicative cost.

Minimizing communication costs is of utmost importance. Human processing costs are Titeratively requesting information, updating our beliefs, and repeating until a satisfactory solution is reached. One approach is Bayesian optimization, which learns to maximize an unknown objective function, e.g. $f(x) = \langle c, x\rangle$, in as few evaluations of f as possible. A key assumption in Bayesian optimization is that f is a black box with unknown structure. Grey-box Bayesian optimization methods relax this assumption, improving sample complexity by taking advantage of structure in the optimization problem (Astudillo & Frazier, 2022).

Two examples of this are when the objective consists of composite functions (Astudillo & Frazier, 2019) and multi-fidelity Bayesian optimization (Poloczek et al., 2016; Foumani et al., 2023). Preferential Bayesian optimization (Astudillo et al., 2023). (move to related work)

We extend grey-box Bayesian optimization by considering a new setting: Agents can gather information about problem substructure directly, without assuming compositional structure.

We then extend grey-box Bayesian optimization to the dialogue setting by proposing a query language through which agents can request and receive information about the unknown optimization problem. The query language is designed to be described concisely in natural language.

We propose a method, Language to Symbolic Optimization (LSO) that collaborates with humans by requests information incrementally, decreasing human communication costs. LSO parses natural language responses to logical forms, which are used to inform a symbolic optimization algorithm that plans what to say next.

We evaluate LSO on DialOp, a set of 3 decision-oriented dialogue tasks (Lin et al., 2023).

2. Related work

3. Decision-oriented dialogue

The goal of decision-oriented dialogue (DOD) is to solve an optimization problem, such as

$$\begin{array}{ll} \text{maximize} & \langle w, x\rangle \\ \text{subject to} & x \in C. \end{array} \tag{1}$$

Solving this problem is straightforward when both w and C are known, and the problem size (dimension $\dim(w)$ and number of constraints $|C|$) is not too large.

We consider the setting where the parameters w, constraints C, and decision variables x of the optimization problem are partitioned between dialogue participants, requiring them to exchange information (about w and F)[1] and make decisions (x) collaboratively. We take the perspective of one dialogue participant, who must communicate with other participants

[1] Cornell Tech [2] Columbia University. Correspondence to: Justin T. Chiu <chiu.justin.t@gmail.com>.

Proceedings of the 40^{th} International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202, 2023. Copyright 2023 by the author(s).

[1] Preference or reward learning is the setting where w is unknown. Constraint acquisition is where C is unknown.

to solve the problem.

Describe prior and observation model. We assume a fully factored prior over w, $p(w) = \prod_i p(w_i)$ with $w_i \sim N(\mu_0, 1)$. The response model, $p(y|a, w)$, models responses y to actions a given parameters w. The action space is extended to $a \in \mathcal{X} \cup \mathcal{Q}$, the union of the decision space and query space. Responses are yes/no or categorical.

4. Planning

We first focus on the setting where the parameters $w \in \mathbb{R}^n$ are unknown, ignoring constraints C. Our goal is to choose the most informative and utility-aware action at every turn in order to select a good terminal $x \in \mathcal{X} = \{0, 1\}^n$.

We plan by optimizing the knowledge gradient acquisition function (Frazier & Powell, 2007). The knowledge gradient is an acquisition function with one-step lookahead. It chooses the next action that gathers information which results in the largest improvement of a subsequent decision. The subsequent decision is not restricted to previously observed values, which is especially important in our setting where actions may not be in the decision space $\mathcal{X}$. Formally, the knowledge gradient acquisition function is given by the following:

$$\begin{aligned} \operatorname*{argmax}_a \mathbb{E}_{y|h,a} \left[\max_x \mu_{w|h,a,y}(x) \right] \\ - \max_x \mu_{w|h}(x), \end{aligned} \tag{2}$$

where $\mu_w(x) = \mathbb{E}_w [\langle w, x \rangle]$, the mean reward. Costs: subtract (textbook) or divide (cost-aware multi-fidelity BO)?

The acquisition function in equation 2 has an argmax over actions $a \in \mathcal{X} \cup \mathcal{Q}$. In the worst case this means that we must solve the inner problem $\max_x \mu_{w|a,y}(x)$ for each action a and response y, which requires calling a solver each time. How to prevent computational blowup? Common to sample y, but that isn't enough. x is also discrete. Is this easy in weighted linear sum assignment problems?

5. Query language

This section discusses the query language $\mathcal{Q}$ and observation model. Bradley-Terry and Plackett-Luce for pairwise and ranking comparisons respectively.

$$p(x \succ y|w) = \frac{e^{w_x}}{e^{w_x} + e^{w_y}} \tag{3}$$

6. Questions

1. Is there a way to not call the solver $|a| * |y|$ times? Some problem-dependent linear optimization trick?

2. Is there a principled way of assigning cost to actions?

   Learned from static data, since we can estimate the utility of any given action? Can be a very primitive kind of learning, e.g. linear function of size(a).

are there other acquisition functions that are 1-step lookahead but aren't intractable like KG? can still do EI w/ 1-step lookahead. (pointers to lit would be helpful!) lit: http://proceedings.mlr.press/v124/lee20a/lee20a.pdf

References

Astudillo, R. and Frazier, P. I. Bayesian optimization of composite functions, 2019.

Astudillo, R. and Frazier, P. I. Thinking inside the box: A tutorial on grey-box bayesian optimization. *CoRR*, abs/2201.00272, 2022. URL https://arxiv.org/abs/2201.00272.

Astudillo, R., Lin, Z. J., Bakshy, E., and Frazier, P. I. qeubo: A decision-theoretic acquisition function for preferential bayesian optimization, 2023.

Foumani, Z. Z., Shishehbor, M., Yousefpour, A., and Bostanabad, R. Multi-fidelity cost-aware bayesian optimization. *Computer Methods in Applied Mechanics and Engineering*, 407:115937, mar 2023. doi: 10.1016/j.cma.2023.115937. URL https://doi.org/10.1016%2Fj.cma.2023.115937.

Frazier, P. and Powell, W. The knowledge gradient policy for offline learning with independent normal rewards. In *2007 IEEE International Symposium on Approximate Dynamic Programming and Reinforcement Learning*, pp. 143–150, 2007. doi: 10.1109/ADPRL.2007.368181.

Lin, J., Tomlin, N., Andreas, J., and Eisner, J. Decision-oriented dialogue for human-ai collaboration. *arXiv preprint arXiv:2305.20076*, 2023.

Neiswanger, W., Yu, L., Zhao, S., Meng, C., and Ermon, S. Generalizing bayesian optimization with decision-theoretic entropies. In Oh, A. H., Agarwal, A., Belgrave, D., and Cho, K. (eds.), *Advances in Neural Information Processing Systems*, 2022. URL https://openreview.net/forum?id=tmUGnBjchSC.

Poloczek, M., Wang, J., and Frazier, P. I. Multi-information source optimization, 2016.

Acknowledgements

People

A. EHIG

Recent work has presented a unified perspective on BOpt to decision-theoretic entropies, proposing a framework that allows for the principled derivation of acquisition functions for custom tasks (Neiswanger et al., 2022). This is referred to as the expected $H_{l,A}$-information gain (EHIG), where $H_{l,A}$ is a utility l and action A aware generalization of the Shannon entropy. While this allows for unifying entropy search and BOpt, not sure if we can do anything with it yet.