Differentiable User Models

Alex Hämäläinen[1] **Mustafa Mert Çelikok**[1] **Samuel Kaski**[1,2]

[1]Department of Computer Science, Aalto University
[2]Department of Computer Science, University of Manchester

Abstract

Probabilistic user modeling is essential for building machine learning systems in the ubiquitous cases with humans in the loop. However, modern advanced user models, often designed as cognitive behavior simulators, are incompatible with modern machine learning pipelines and computationally prohibitive for most practical applications. We address this problem by introducing widely-applicable differentiable surrogates for bypassing this computational bottleneck; the surrogates enable computationally efficient inference with modern cognitive models. We show experimentally that modeling capabilities comparable to the only available solution, existing likelihood-free inference methods, are achievable with a computational cost suitable for online applications. Finally, we demonstrate how AI-assistants can now use cognitive models for online interaction in a menu-search task, which has so far required hours of computation during interaction.

1 INTRODUCTION

User modeling constructs informative representations of individual users to enable computational systems to customize and adapt their behavior for them [Li and Zhao, 2020]. It has been extensively studied over the years, also recently in recommender systems [Yu et al., 2019, Yuan et al., 2020], human-in-the-loop machine learning [Daee et al., 2018] and AI-assistants [Horvitz et al., 2013, Dafoe et al., 2021]. Machine learning is needed in user modeling to infer user-specific information based on observed user behavior. Depending on the application, the inferred information can be the end result or, for instance, used to parameterize a user simulator to form predictions of user behaviors to guide the behavior of the system.

Traditionally, salient use cases of user modeling, such as recommendation engines, have utilized user-specific preference profiles based on users' history. However, these approaches are not sufficiently powerful in more complex interactive applications, where the user plans and interacts strategically, for instance in human-AI collaboration and human-in-the-loop decision-making. On the other hand, while recent ML research has shown significant success in learning accurate neural models directly from data, this is an infeasible approach in user modeling in general, as typical user-driven applications are data starved. In contrast, recent approaches, e.g., Kangasrääsiö et al. [2019], Moon et al. [2022], have utilized advanced general-purpose behavioral models, based on cognitive science, in a Bayesian setting and received encouraging results with limited data. The probabilistic treatment of the problem enables taking uncertainty of the inferences into account — which fundamentally allows the system to balance between exploration and exploitation in interaction with the user.

A prominent body of these advanced cognitive models is based on computational rationality [Lewis et al., 2014, Gershman et al., 2015], which posits that seemingly irrational behaviors of humans are rational under their cognitive bounds. It follows that human behaviors can be accurately modeled as a result of RL-based optimization given that the underlying decision-theoretic framework and the optimization procedure are specified such that they accurately capture the appropriate bounds governing human cognition, such as the limits of computational capacity. A concrete example of such a model, which we consider in our experiments, is the model of rational menu search [Chen et al., 2015]. This cognitive model describes human search behavior in terms of eye movements when searching for a target item in a computer dropdown menu, while encoding the limitations of human cognition and perception when processing visual information.

Despite their many benefits, these advanced cognitive models are currently not used beyond small-scale practical applications due to two important factors: (1) they are expressed

Accepted for the 39th Conference on Uncertainty in Artificial Intelligence (UAI 2023).

as non-differentiable simulators and hence incompatible with modern machine learning frameworks and (2) they are computationally infeasible to be used directly in realistic applications.

In this paper, we address these limitations: we enable computationally efficient probabilistic user modeling suitable for real-time applications — even with advanced cognitive models that lack a closed-form likelihood. We do this by combining the best of gradient-based and Bayesian learning: we show how one can develop differentiable user models which are sample-efficient by leveraging prior knowledge from non-differentiable cognitive models and can quantify uncertainty in their estimates. As a result, the surrogates become widely applicable with online computational cost independent of the complexity of the original models. The contributions of this work are:

- We introduce a way of enabling computationally efficient inference with cognitive user models by building generalizable differentiable surrogates for them through meta-learning.
- We demonstrate a flexible way of leveraging any existing user data during surrogate training to address possible model misspecification in cognitive models, especially in the case of action noise.
- With neural processes as example surrogate models, we demonstrate comparable user modeling accuracy to current methods with a computational time suitable for online applications.

Our work removes a key computational bottleneck currently hindering incorporation of users into probabilistic programming models. Probabilistic user modeling based on cognitive models can now be applied widely without extensive computational budgets.

2 DIFFERENTIABLE USER MODELS

This work considers computationally efficient probabilistic user modeling for interactive settings, between a user from a user population and a computational system. User modeling is brought in to guide adaptation of the behavior of such a system for individual users; user modeling is needed for (i) inferring user-specific information from observed user behaviors and (ii) then using the information for user behavior prediction. The following subsections detail the specifics of current approaches and their limitations, together with our approach for addressing these limitations.

2.1 PROBABILISTIC USER MODELING WITH COGNITIVE MODELS

Following the intuition presented by Kangasrääsiö et al. [2019], we formulate the probabilistic user modeling setting as follows: a population of users is engaged with a distribution of decision-making tasks denoted by $p(\theta_T)$. Each modeling scenario involves a user $\theta_U \sim p(\theta_U)$ and is fully described by the respective user and task specific parameters $\theta = \{\theta_T, \theta_U\}$. The users are assumed to generate their policies π through an implicit process $\mathcal{P}_\theta$ which they execute to generate pairs of states and actions $(\mathbf{s}, \mathbf{a})$. The system has access to a cognitive model $p(\pi \mid \theta)$ approximating the true process and a corresponding prior $p(\theta)$. An important task corresponding to this user modeling setting is the inference problem of approximating the posterior

$$p(\theta \mid (\mathbf{s}, \mathbf{a})) \propto p((\mathbf{s}, \mathbf{a}) \mid \pi) p(\pi \mid \theta) p(\theta). \tag{1}$$

Computing the posterior $p(\theta \mid (\mathbf{s}, \mathbf{a}))$ and then using the likelihood model $p(\pi \mid \theta)$ for computing the corresponding posterior predictive distribution $p(\pi \mid (\mathbf{s}, \mathbf{a})) = \int p(\pi \mid \theta) p(\theta \mid (\mathbf{s}, \mathbf{a})) d\theta$ would be the Bayesian choices for achieving the objectives (i) and (ii).

For cognitive models, the likelihood $p(\pi \mid \theta)$, required for solving the Bayesian inference task for the posterior, is typically not evaluable in closed-form due to the simulator-type nature of these models. So far, this issue has been circumvented by utilizing exclusively likelihood-free inference (LFI) methods, such as approximate Bayesian computation (ABC) [Sisson et al., 2018, Sunnåker et al., 2013] and Bayesian optimization for likelihood-free inference (BOLFI) [Gutmann et al., 2016], as proposed by Kangasrääsiö et al. [2019] and Moon et al. [2022]. The basic idea of LFI is to replace the computationally expensive simulator $p(\pi \mid \theta)$ with an approximation that is separately learned on the observed data [Gutmann et al., 2016], in this case the $(\mathbf{s}, \mathbf{a})$ from each user. For user modeling, this approach has two problems: this process requires numerous computationally expensive simulations with the cognitive model and the data in typical user modeling applications often is too scarce for learning a new model independently for each user.

Moon et al. [2022] proposed circumventing the computational complexity by learning a generalizable policy-modulation network as a surrogate for the original model, i.e. $p(\pi \mid \theta)$, and obtained significant speed-ups for inference. However, as noted by the authors, their approach is still prohibited by the computational cost of LFI needed for approximating the posterior, and is not feasible for real-time inference. Similarly, the simulation costs of cognitive models $p(\pi \mid \theta)$ are often too expensive to enable estimating the posterior predictive in real-time applications, even if the posterior was readily available. Furthermore, even though LFI methods are developing fast, practical interactive settings may require hierarchical approaches for generalizing across the user models, which has been traditionally difficult with LFI [Turner and Van Zandt, 2014]. An additional problem with LFI-based modeling is the sensitivity to model misspecification, which is very likely in user models.

2.2 AMORTIZATION FOR COGNITIVE MODELS

In this work, we seek to address the limitations of current approaches and to enable efficient computation for both the likelihood and posterior models so that the posterior predictive distribution is practical to approximate. Our approach is to amortize posterior predictive inference through surrogate modeling. Training generalizable surrogates offline enables using them during online interaction without extensive computation.

While this approach would solve the issue of online computational complexity, the offline complexity of simulating sufficient amounts of training data for them will still be an issue due to the vast diversity of different behaviors the cognitive models are able to express. In particular, even if training a generalizable surrogate for a cognitive simulator would be achievable, as done in the work of Moon et al. [2022], training a surrogate directly for approximating the posterior can be ultimately be computationally intractable if constructing the training data requires numerous repeated evaluations with LFI (for the reference, Kangasrääsiö et al. [2019] reported that even a single LFI result would require at least 700 CPU hours with the menu search model). Furthermore, as we will later discuss in Section 2.5, data-efficiency in training the surrogates is also otherwise a desirable factor as it helps combating model misspecification in cognitive models.

2.3 CASTING SIMULATOR-BASED MODELING AS META-LEARNING

In order to make the surrogate training more sample-efficient, we approach amortization task from meta-learning perspective. Here, the key insight is that both the likelihood and posterior models can be learned jointly with an appropriate policy approximation task, without ever needing to approximate the true posterior $p(\theta \mid (\mathbf{s}, \mathbf{a}))$, if one is satisfied with using a latent representation $z \in \mathcal{Z}$ to capture user-specific information. Following this intuition, we generalize the likelihood and posterior models to mappings h and g:

Definition 2.1 (Amortization for cognitive models)**.** Let $\mathcal{S}$ and $\mathcal{A}$ denote the state and action spaces corresponding to the user model and $\mathcal{O} = \bigcup_n (\mathcal{S} \times \mathcal{A})^n$ be a collection of m observations over behavior generated by an individual user θ_U in a task θ_T. Amortization for cognitive models corresponds learning the following functions such that they are evaluable during online interaction:

1. Inference of user and task representations, done by the mapping $h: \mathcal{O} \rightarrow P(\mathcal{Z})$, where $P(\mathcal{Z})$ denotes a probability distribution over a joint user and task representation space $\mathcal{Z}$, which aims to capture the properties governing user behavior.
2. User behavior prediction, done by the mapping $g: \mathcal{S} \times \mathcal{Z} \rightarrow P(\mathcal{A})$, where $P(\mathcal{A})$ is a probability distribution over user action space.

In line with the Definition 2.1, we amortize the computation for the posterior predictive distribution over a cognitive model through learning generalizable surrogates for the mappings h and g. Intuitively, we are here building on the conceptual similarity between Bayesian methods and meta-learning (previously discussed, e.g., by Grant et al. [2018] and Garnelo et al. [2018b]), and consider the mapping h, i.e., computing the posterior over the user representation as equivalent to task-specific adaptation during meta-testing and the mapping g, i.e., computing the likelihood as analogous to prediction.

To formalize the idea, let $s \in \mathcal{S}$, $a \in \mathcal{A}$ and $(\mathbf{s}, \mathbf{a}) = \{(s_1, a_1), \ldots, (s_n, a_n)\} \in \mathcal{O}$. Our goal is to learn the mappings h and g, with optimizable parameters $\{\psi, \phi\}$, to approximate the posterior $p_\psi(z \mid (\mathbf{s}, \mathbf{a}))$ and the likelihood $p_\phi(a \mid s, z)$ with respect to a latent representation $z \in \mathcal{Z}$. The corresponding posterior predictive model can here be written as $p_{\{\psi,\phi\}}(a \mid s, (\mathbf{s}, \mathbf{a})) = \int p_\phi(a \mid s, z) p_\psi(z \mid (\mathbf{s}, \mathbf{a})) dz$. The surrogates should jointly minimize the following objective for policy approximation, while generalizing over the ground-truth user and task population ($\theta \sim p(\theta)$ and $\pi \sim p(\pi \mid \theta)$):

$$\min_{\phi,\psi} \mathbb{E}_{\theta \sim p(\theta), s \in \mathcal{S}} \Big[\delta \big[\pi(a \mid s), p_{\{\psi,\phi\}}(a \mid s, (\mathbf{s}, \mathbf{a})) \big] \Big], \quad (2)$$

where δ is a dissimilarity function (e.g., KL-divergence) and the observations $(\mathbf{s}, \mathbf{a})$ are assumed to have been generated by executing π in the underlying environment. In Section 3, we demonstrate how a solution to this problem can be approximated with neural processes.

Consistently with numerous current meta-learning approaches (e.g., Finn et al. [2017], Garnelo et al. [2018b]), we propose a modeling workflow with separate offline (meta-training) and online (meta-testing) phases described below. We additionally expand on mitigating the effects of possible model misspecification in cognitive models.

2.4 META-TRAINING AND META-TESTING

Algorithm 1 specifies the proposed meta-training procedure, to enable generalization of the surrogates h and g over the population of interest $p(\theta)$. The procedure needs to be complemented with an appropriate meta-learning loss for approximating a solution to Eqn. 2 in terms of $\{\psi, \phi\}$, depending on the implementations of h and g. In Section 3, we exemplify this with neural processes.

The corresponding meta-testing, i.e., task-specific adaptation phase is straightforward: mappings h and g can be utilized for inferring user representations $z \sim p_\psi(z \mid (\mathbf{s}, \mathbf{a}))$

w.r.t. observed $(\mathbf{s}, \mathbf{a})$ and for predicting user behaviors $a \sim p_\phi(a \mid s, z)$ on states of interest $s \in \mathcal{S}$.

Algorithm 1 Meta-training cognitive model surrogates

Require: A distribution over users: $p(\theta_U)$
Require: A distribution over tasks: $p(\theta_T)$
Require: A cognitive model: $p(\pi \mid \theta)$
Initialize h and g with $\{\psi, \phi\}$
repeat
  Sample $\theta = \{\theta_U, \theta_T\}$, $\theta_U \sim p(\theta_U)$, $\theta_T \sim p(\theta_T)$
  Generate $\pi \sim p(\pi \mid \theta)$
  Generate n trajectories $(\mathbf{s}, \mathbf{a})$ by executing π
  Optimize $\{\psi, \phi\}$ with respect to $(\mathbf{s}, \mathbf{a})$ with an appropriate training loss
until done

Note that consistently with Garnelo et al. [2018b], the proposed meta-learning workflow deliberately differs from many other popular meta-learning approaches, such as model-agnostic meta-learning (MAML) [Finn et al., 2017] and Reptile [Nichol and Schulman, 2018], by fully excluding the gradient-based optimization loop during task-specific adaptation phase. Instead, the adaptation phase is here reduced to a forward pass through h. Not only is this computationally faster, enabling online computation, the probabilistic nature of our approach can also enable interactive systems to balance between exploration-exploitation trade-offs. As we demonstrate in our experiments, these benefits additionally translate into improved modeling accuracy.

2.5 MODEL MISSPECIFICATION IN COGNITIVE MODELS.

Model misspecification is a relevant issue in behavioral user modeling. While typical LFI-approaches are highly sensitive to misspecification, this can be mitigated with our approach by combining observed user data with simulated data and meta-training the surrogates again, when new observations become available. We demonstrate in Section 4.2 that this approach enables balancing between modeling accuracy and data requirements — especially in practical interactive user modeling applications which only have limited collections of user behavior datasets available.

3 USER MODELING WITH NEURAL PROCESSES

We use neural processes (NP) [Garnelo et al., 2018b] as an example solution for implementing and learning the mappings h and g of Definition 2.1. First, we briefly cover the relevant background on NPs and then explain in detail how they can be adapted for user modeling.

3.1 BACKGROUND ON NEURAL PROCESSES

Neural processes [Garnelo et al., 2018b] are a family of neural latent-variable models combining properties of neural networks and Gaussian processes (GP). Specifically, they are differentiable solutions for representing uncertainty over functions that may be utilized for few-shot approximation. For our purposes, NPs are particularly fitting as they match Definition 2.1 and that the meta-learning objective (Eqn. 2) can be readily computed for them.

NPs model a set of functions $\{f_d\}_d$ where each $f_d : X \to Y$ is assumed to be drawn from an underlying stochastic process $f_d \sim F$. NP approximates the underlying process F with a neural network g. As each function f_d drawn from the process F represents an individual instantiation of the process, a latent variable z is introduced for capturing the instance-dependent variation in F as $f_d(x) = g(x, z)$. NPs consist of an encoder, an aggregator and a conditional decoder. The encoder is a neural network for constructing representations $r_i = h_\phi((x, y)_i)$ at given observations $(x, y)_i$. The aggregator, α, constructs permutation-invariant summaries of the encoded representations as $r = \alpha(\{r_i\}) = \frac{1}{n}\sum_{i=1}^{n} r_i$. The summaries are further utilized to parametrize a (multivariate Gaussian) latent distribution $z \sim \mathcal{N}(\mu(r), I\sigma(r))$. The conditional decoder, $g_\psi(x_T, z)$, is a neural network that is conditioned on samples from the latent distribution to estimate $f_d(x_T) = y$ at locations x_T.

NP meta-training procedure samples individual instantiations $f_d \sim F$ of the stochastic process F. Here, each function f_d is evaluated at a varying number of inputs to produce a dataset of tuples $(x, y)_i^d$. Each dataset is then divided into separate *context* $(x_{1:m}, y_{1:m})$ and *target* $(x_{m+1:n}, y_{m+1:n})$ sets. Intuitively, here the context set represents the fully observed function evaluations while the target $x_{m+1:n}$ represents the locations at which the model aims to approximate $y_{m+1:n}$. The context and target sets are input to the encoder and the conditional decoder respectively, and the model parameters $\{\phi, \psi\}$ are optimized with respect to the NP-variant of Evidence lower-bound (ELBO). For further information about NPs and their training, see Garnelo et al. [2018b]. Finally, note that the low complexity of NPs ($\mathcal{O}(n + m)$) makes them suitable for real-time scenarios.

3.2 ADAPTING NEURAL PROCESSES FOR USER MODELING

Neural processes can be adapted as concrete implementations for the required mappings h and g and for approximating a solution for Equation 2 within the proposed meta-training procedure (Algorithm 1). First, we recognize that Equation 2 is essentially a function approximation problem to which NPs can be applied — the true behavior-generative process $p(\pi \mid \theta)$ can essentially be treated as a stochastic

process $\mathcal{P}$ where each instantiation $\pi \sim \mathcal{P}$ represents a policy. The NP latent variable z is utilized for capturing user/task representations and the mappings h and g can be implemented with the NP encoder $p_\phi(\pi \mid z)$ and conditional decoder $p_\psi(z \mid (\mathbf{s}, \mathbf{a}))$. The meta-training procedure is adapted as follows: the sampled behaviors $(\mathbf{s}, \mathbf{a})$ are split into context and target sets and the parameters $\{\psi, \phi\}$ can be optimized according to NP-ELBO.

In addition to the vanilla NPs, we consider also attentive neural processes (ANP) Kim et al. [2019], conditional neural processes (CNP) [Garnelo et al., 2018a] and attentive conditional neural processes (ACNP). ANPs are essentially NPs with the difference of including attention in the encoder architecture. The attention acts as a local latent variable, allowing ANPs to capture both global and local information affecting user behaviors. CNPs (and ACNPs) implement the latent encoding h as a deterministic mapping, thus lacking an important ability of sampling on $\mathcal{Z}$.

4 EXPERIMENTS

We conduct three experiments where we compare our approach against other ways one could conceivably try to solve the problem — although this has not been previously done. The first is a demonstration in a benchmark gridworld environment. The second is a menu search task where a cognitive user model, justified and validated by earlier cognitive science studies, allows us to study real-user performance with simulations. The third experiment is a reasonably realistic menu search assistant scenario.

Comparison methods and baselines. We assess the modeling capabilities of the proposed solution in terms of its ability to predict the actions of individual agents, as a function of the number of previous observations of their behavior in the modeling task of Equation 2. This metric directly evaluates the posterior predictive but also indirectly the quality of the posteriors over user representations $z \in \mathcal{Z}$. Unless otherwise specified, the experiments aim to simulate realistic user modeling applications by limiting the training data to observations from ~ 1000 simulated users.

We compare our approach against two baselines and three alternative surrogate architectures. The alternative surrogates are transformers trained with MAML and Reptile, and a standard transformer. MAML and Reptile act as alternative representative meta-learning approaches to the policy approximation task over user population, while the transformer intends to provide a reference point for the performance of sequential models which are frequently used in alternative user modeling domains, such as sequential recommendation. None of the alternative surrogate architectures are fully consistent with the proposed meta-learning procedure and are applied to the policy approximation task on simulated data directly instead. We also include comparisons between

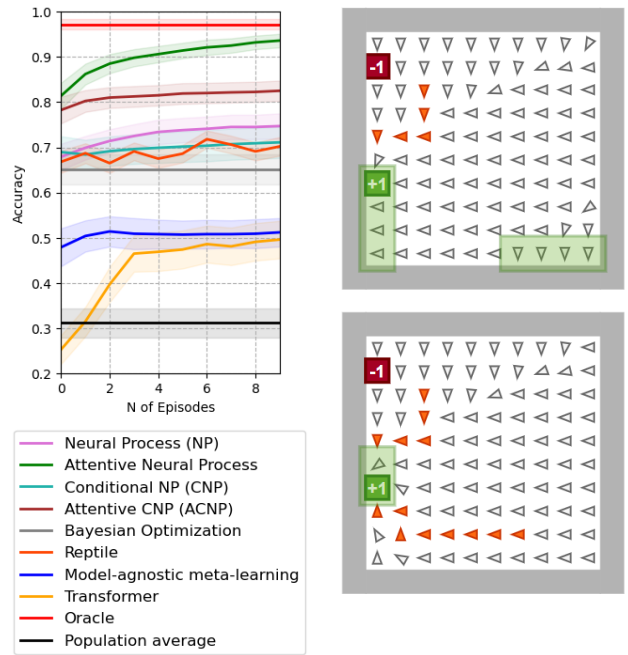

Figure 1: Gridworld results. **Left:** Modeling accuracy as a function of the number of observed full episodes. The best NP-based model (here ANP) achieves comparable results to the upper bound given by an oracle; all NP-based models are clearly better than alternatives. The figure illustrates the gradual improvement of the predictions of NPs as more episodes are perceived. The BO results are averaged over the number of context trajectories due to the small sample size. **Right:** Illustration of ANP uncertainty update on policy predictions. The predictions (gray arrows) align towards the implicitly inferred possible goal states (green rectangles). In the upper figure, the predictions are conditioned on one observed trajectory (orange arrows). In the lower figure, we observe that the system implicitly infers the location of the positive reward, within the accuracy of two squares, after perceiving the second trajectory. The dark green and red squares are the true positive and negative reward states.

several alternative NP architectures. Details are included in the Supplement.

The two baselines are a Bayesian Optimization (BO) model and a population average predictor. Furthermore, we provide results from an oracle, acting as an upper bound for the performance of any solution, including LFI. Both the BO baseline and the oracle utilize the cognitive model $p(\pi \mid \theta)$ directly for prediction — the oracle parametrizes the cognitive model with the true user parameters while BO utilizes MAP estimates. The population average predictor directly approximates the population level action distribution $p(a \mid s)$ for each state without any user-specific conditioning.

Even though it would be an important baseline, we were not able to produce any representative results with LFI due

Table 1: Modeling accuracy as a function of the number of observed full episodes in the menu-search setting of Section 4.2.

Episodes	ANP	Reptile	MAML	Transformer	Oracle	Population avg.
0	$\mathbf{0.937 \pm 0.011}$	0.829 ± 0.054	0.774 ± 0.033	0.920 ± 0.035	$\mathbf{0.970 \pm 0.002}$	0.921 ± 0.012
1	$\mathbf{0.953 \pm 0.011}$	0.921 ± 0.021	0.916 ± 0.026	0.922 ± 0.021	...	...
2	$\mathbf{0.954 \pm 0.011}$	0.930 ± 0.020	0.928 ± 0.025	0.931 ± 0.017	...	...
5	$\mathbf{0.955 \pm 0.010}$	0.944 ± 0.017	0.943 ± 0.021	0.926 ± 0.012	...	...
9	$\mathbf{0.955 \pm 0.010}$	0.954 ± 0.016	0.952 ± 0.014	0.928 ± 0.009	...	...

to its immense computational complexity. For reference, Kangasrääsiö et al. [2019] compared several LFI-methods, including BO, on exactly the same Menu Search model used in our experiments. They gave 700h of CPU-time for each method to run only one individual inference task and noted that it is likely that none of the methods converged. Obtaining conclusive accuracy results with LFI in our experiment setting is practically intractable as at least hundreds of individual inference results would be needed. Here, we consider the BO-baseline as an approximate lower-bound for LFI performance. Although converging faster than LFI, BO is still computationally very heavy, due to the expensive simulation costs, and feasible only in our first experiment.

4.1 EXPERIMENT 1: GRIDWORLD ENVIRONMENT

Setting. The first experiment scenario is based on a simple 10×10 gridworld environment. In this setting, we consider modeling Monte Carlo Tree Search (MCTS) [Browne et al., 2012] agents with unknown reward functions and MCTS parameters. This benchmark scenario evaluates the modeling system's ability to approximate the uncertainty over user policies. In this experiment, we assume that a generative user model and a parameter prior are available, capturing the true generative process of the population.

The gridworld environment is defined as a partially observable Markov decision process (POMDP) with deterministic transition dynamics. The action space consists of four possible actions that correspond to the agent relocating from its current state to adjacent states. Each gridworld scenario always contains two reward states - one with a positive reward and one with a negative reward. The agents gain no rewards or penalties other than from the given states. The full setting details are given in the Supplement.

Modeling task. The modeling task is to predict the subsequent actions of agents sampled from the population. Each scenario assigns the modeling system with observed trajectories from a varying number of previous episodes and a partial trajectory from the current episode generated by the agent. The task is to predict the remaining actions of the trajectory in the current episode. All information about the agent, excluding the observed trajectories, is hidden during both training and evaluation (except for the oracle).

Results. The NP-family models are mostly able to outperform all the baselines (Fig. 1) with the ANP converging close to the performance of the oracle. It is likely that the BO-baseline has not properly converged, although given clearly the largest amount of computation time, and it may not act as a reliable approximation for LFI performance. Finally, the transformer and MAML are unable to generalize to the task, likely due to the too limited amount of training data.

Comparisons between the NP-family models suggest that, in terms of NP architecture, the most impactful factor contributing to the modeling performance is attention, i.e., local latent variables, as ANP and ACNP outperform their non-attentive counterparts. Consistent with the probabilistic treatment of h (Def. 2.1), stochasticity of the global latent variables z (ANP and NP) also seems to improve the results. Because ANP was clearly the best of the NP methods, and hence remaining NP-models would not affect conclusions, we omit the NP, CNP and ACNP models for the following experiments, to save computation.

4.2 EXPERIMENT 2: MENU SEARCH ENVIRONMENT

Setting. Our second experiment is based on the Menu Search model of Kangasrääsiö et al. [2019], a modified version of Chen et al. [2015]. The Menu Search model is a cognitive model describing human search behavior in terms of eye movements (saccades) when searching for a target item in a computer dropdown menu. Motivated by *computational rationality* [Gershman et al., 2015], the model simulates user behavior as a result of optimizing the search behavior with RL given their cognitive constraints of the user. The details are given in the Supplement.

Modeling task. In this experiment, we apply our method for modeling agents whose search behavior is specified by the Menu Search model. As in the first experiment, we train the model parameters on data simulated with the given cognitive model. For each simulation, we sample a new menu, together with its element-wise information about the target word, as specified by Kangasrääsiö et al. [2019].

Modeling accuracy. Table 1 summarizes the obtained modeling accuracies. We notice that after one observed tra-

Table 2: Modeling accuracies for different numbers of observed full episodes with the ANP-based system when trained with data partially from a misspecified user model and partially from the true population. Here, the percentages denote the share of the training data generated with the misspecified user model.

Episodes	ANP 0%	ANP 25%	ANP 50%	ANP 75%	ANP 100%
0	$\mathbf{0.937 \pm 0.011}$	0.920 ± 0.015	0.895 ± 0.016	0.888 ± 0.013	0.852 ± 0.017
1	$\mathbf{0.953 \pm 0.011}$	0.923 ± 0.012	0.899 ± 0.014	0.891 ± 0.013	0.854 ± 0.016
2	$\mathbf{0.954 \pm 0.011}$	0.925 ± 0.012	0.901 ± 0.014	0.892 ± 0.012	0.857 ± 0.016
5	$\mathbf{0.955 \pm 0.010}$	0.926 ± 0.012	0.902 ± 0.014	0.894 ± 0.012	0.862 ± 0.016
9	$\mathbf{0.955 \pm 0.010}$	0.926 ± 0.012	0.902 ± 0.014	0.895 ± 0.012	0.865 ± 0.016

jectory, the ANP-based model achieves results comparable to the oracle upper bound. Unlike in the previous experiment, most of the users seemed to converge to a relatively narrow and finite set of search strategies, simplifying the difficulty of the modeling problem. As a result, MAML and the transformer achieve clearly higher relative accuracy than in the previous experiment, despite the limited training data.

Model misspecification. We study the effects of model misspecification in cognitive models by repeating the modeling task with a noisy model. This model represents an otherwise accurate Menu Search model, but roughly 35% of the saccades are modeled randomly into incorrect locations instead of following the policy of the correct model (full implementation details in the Supplement). We repeat the meta-training with different percentages of data obtained from the true user population. We explore both our solution's robustness against the model with action noise and its ability to adapt to the true generative process.

The results are gathered in Table 2. First, we observe that our approach can remain robust against user model noise: even when trained solely on data coming from the noisy model (ANP 100%), the modeling accuracy remains reasonably good and surpasses the accuracy of the noisy model ($\approx 65\%$). Secondly, it can be seen how our solution adapts to the ground-truth generative process when the proportion of the ground-truth data increases. We repeated the scenario by meta-training the ANP only on data from the true user population. We found that the noisy model improved the results when the number of observed real users was under 200 (i.e., here percentage $< 20\%$), after which it had a hindering effect on the predictions. However, we expect that the utility of misspecified models can be significantly higher in more complex modeling problems where more data is required to generalize to the problem.

4.3 EXPERIMENT 3: MENU SEARCH ASSISTANT

Setting. In our third experiment, we aim to demonstrate the practical utility of the proposed approach for interactive systems by extending the Menu Search environment into a reasonably realistic AI-assistant scenario. First, we scale the environment to consider two levels of hierarchy: the menu consists of a main menu whose elements act as links to sub-menus; we use the menus of the previous experiment.

Secondly, we introduce an AI assistant equipped with the proposed user modeling system. The task of the assistant is to utilize the modeling system to infer the target elements of the users based on observed search behaviors in the current menu, and to propose sub-menus for the users. Intuitively, a successful assistant should guide the users to menus that are likely to contain the true target for them, to shorten their search time. The assistant is allowed to provide any guidance only after the user is independently searched through at least one sub-menu. We implement the assistant as a simple rule-based agent that conditions its actions on the simulated user behaviors $a \sim p_\phi(a \mid s, z)$, $z \sim p_\psi(z \mid (\mathbf{s}, \mathbf{a}))$. Further details on the experiment setting are in the Supplement.

Results. Table 3 compares the performance of the resulting assistant against a non-assisted user, a MAML-based, a Reptile-based, and a transformer-based solutions. The MAML and Reptile-based solutions require gradient-computation during test-time leading to modeling times greatly higher than the response time between human actions ($\approx$ 300ms) in this experiment. This prevents online user model updates, hence hindering the effectiveness of the assistance. We also include results with an assistant that has full knowledge of the users' target elements to provide an upper-bound. We notice that the ANP-guided assistant can significantly reduce the user's search time and almost reaches the upper-bound performance of the assistant that has perfect knowledge. The observed results are encouraging regarding the ability of our solution to guide the behaviors of real-time interactive systems.

5 RELATED WORK

Our work connects to a larger body of research considering user modeling in interactive AI. For instance, Carroll et al. [2019] and Strouse et al. [2021] share the idea that efficient interaction with humans requires the AI to have an accurate model of the human. In contrast to many this line of works, our work concentrates on using models based on cognitive and behavioral sciences as priors, instead of ML-experts hand-crafting the models from scratch or learning them

Table 3: User search times and modeling/simulation times per assistant action with different assistant systems in Section 4.3.

Assistant type	Search time (s)	Time saved (%)	Modeling time per action (ms)
No assistance	4.774 ± 0.235	–	–
MAML	4.089 ± 0.645	14.3	1174.922 ± 43.760
Reptile	3.973 ± 0.519	16.8	1053.342 ± 37.988
Transformer	2.918 ± 0.191	38.8	$\mathbf{1.140 \pm 0.442}$
ANP	$\mathbf{2.590 \pm 0.226}$	**45.7**	8.460 ± 6.495
Full knowledge	$\mathbf{2.577 \pm 0.162}$	**46.0**	–

from large collections of user data. Using such models has been impractical up to now, and this the problem we now solve.

Inverse reinforcement learning (IRL) [Ng et al., 2000] considers a related problem to ours, aiming to recover agents' reward functions based on observed behaviors. Although it has been previously utilized also in user-centric problems [Chandramohan et al., 2011], our perspective is more general as we consider inference over arbitrary user parameters (instead of only rewards) and over varying policy-generative algorithms/processes. This allows our approach to be utilized for inference with a wide range cognitive models, where user behaviors are not necessarily optimal and are governed by human biases. Imitation learning (IL) [Hussein et al., 2017], on the other hand, considers learning models to imitate human (expert) behaviors on a given task. The crucial difference to our setting is that, unlike with IL, we do not necessarily seek to solve the task the human is solving, but to probabilistically model humans and their behaviors.

Using transfer and meta-learning in RL problems has been previously widely studied. For instance, Yao et al. [2018] used HiP-MDPs [Doshi-Velez and Konidaris, 2016] for modeling differences in environment dynamics and to further parametrize a policy. Similarly, Galashov et al. [2019] propose a probabilistic framework for sequential decision-making that they instantiate with NPs for meta-learning. In contrast to this line of works, the novelty of our work is not about a generalizable solution to distributions of RL tasks, but rather about a generalizable method for making modeling with cognitive models practical. This is an important distinction because cognitive models are not necessarily compatible with the RL formalism — even when they are, they are based on computational rationality, and specifically tailored to account for cognitive limitations. Adapting these limitations to existing frameworks, such as HiP-MDPs, is not trivial and necessarily requires manual effort.

Our work also connects to a line of research studying inference for decision making agents in the context of probabilistic programming. However, most of the approaches make restricting assumptions either regarding the behavior generative processes of the users or the inference objectives and could be applied only for very limited types of problems. For instance, Zhi-Xuan et al. [2020] consider online inference of boundedly-rational agents but their approach can be applied only in discrete and deterministic environments to capture only agent goals. Furthermore, their solution assumes that the agents start planning their policy from scratch during interaction — in practical interactive settings, humans might already have a partial or complete plans at the beginning of interaction. On the other hand, many other works, such as by Seaman et al. [2018], assume that the likelihood for the generative process $p(\pi \mid \theta)$ can be evaluated for MCMC, which is often an unrealistic assumption with advanced cognitive models.

Many computational approaches motivated by cognitive science share parallels with our objectives. For instance, computational rationality [Lewis et al., 2014, Gershman et al., 2015] and Theory of Mind (ToM) [Premack and Woodruff, 1978] have motivated numerous computational approaches such as Bayesian ToM [Baker et al., 2011], Machine Theory of Mind [Rabinowitz et al., 2018] and the Menu Search model [Chen et al., 2015] for modeling human behaviors. Furthermore, Peltola et al. [2019] utilize ToM for modeling users with their own models of the interactive system in bandit settings. Among others, these models are prime candidates our method can be applied to.

6 DISCUSSION

In this work, we have addressed the so-far unaddressed problem of enabling probabilistic user modeling with complex cognitive models in real-time applications. We introduced a meta-learning approach for training widely applicable differentiable surrogates for approximating posterior predictive estimation with cognitive models. We studied neural process models as example implementations for the surrogates and demonstrated comparable modeling performance to likelihood-free inference with computational cost suitable for online applications. We also showed that the proposed solution allows AI-assistants to utilize cognitive user models computationally feasibly, for instance in a previously studied menu-search task. In a larger scale, the solution not only removes a computational bottleneck currently hindering incorporation of users into probabilistic programming models, but also enables real-time user modeling in various applications where they currently are not possible within usually available computational budgets.

We also demonstrated how the effects of model misspecification in cognitive models can be mitigated in the surrogates by incorporating observed user data in the training. Importantly, we observed that our approach provided robustness against action noise while adapting to the true population as more behavior data became available. Based on these observations, we conclude that the proposed solution can be particularly useful in application domains where user data are limited or behavioral user models can be slightly misspecified, although future studies are still required in settings where the misspecification is caused by more systematic biases.

It is crucial to note that amortization for probabilistic user modeling with cognitive models, as detailed in Section 2, has not been previously widely studied. Apart from LFI, which is computationally intractable for our problems, we are not aware of any solutions which could act as either relevant baselines or alternatives to the proposed approach. Specifically, all the experimented alternative surrogate implementations, such as MAML and transformers, are not fully consistent with the probabilistic nature of the problem, limiting their applicability in practice. We further note that also neural processes feature several compromises in comparison to a fully Bayesian setting with cognitive models: although supporting posterior predictive estimation, they cannot be directly adapted for Bayesian inference in an explicit, predefined parameter space and they do not necessarily follow all constraints coming from the known structure of the behavioral model due to amortization. Future research should adapt and develop alternative surrogate solutions to address these drawbacks.

Interesting avenues for future research include utilizing the surrogates in full probabilistic programming pipelines, although we hypothesize that this should already be possible within certain limits with our approach. Other attractive extensions could consider alternative surrogate architectures to handle, for instance, non-stationarity in cognitive models and settings with multiple data modalities. Regarding ethical considerations, user modeling has always been a double-edged tool and can potentially be abused to serve other interests than those of users — this should be taken carefully into account in all of its applications. As a generic tool to mitigate some of these issues, we recommend combining user systems with privacy preservation with differential privacy.

Acknowledgements

We would like to thank Sebastiaan De Peuter, Pierre-Alexandre Murena and Sammie Katt for their valuable advice and feedback. This work was supported by the Academy of Finland (Flagship programme: Finnish Center for Artificial Intelligence FCAI and decision 345604) Humane-AI-NET and ELISE Networks of Excellence Centres (EU Horizon: 2020 grant agreements 952026 and 951847), and UKRI Turing AI World-Leading Researcher Fellowship (EP/W002973/1). We also acknowledge the computational resources provided by the Aalto Science-IT Project from Computer Science IT.

References

Chris Baker, Rebecca Saxe, and Joshua Tenenbaum. Bayesian theory of mind: Modeling joint belief-desire attribution. In *Proceedings of the annual meeting of the cognitive science society*, volume 33, 2011.

Cameron B Browne, Edward Powley, Daniel Whitehouse, Simon M Lucas, Peter I Cowling, Philipp Rohlfshagen, Stephen Tavener, Diego Perez, Spyridon Samothrakis, and Simon Colton. A survey of monte carlo tree search methods. *IEEE Transactions on Computational Intelligence and AI in games*, 4(1):1–43, 2012.

Micah Carroll, Rohin Shah, Mark K Ho, Tom Griffiths, Sanjit Seshia, Pieter Abbeel, and Anca Dragan. On the utility of learning about humans for human-ai coordination. *Advances in neural information processing systems*, 32, 2019.

Senthilkumar Chandramohan, Matthieu Geist, Fabrice Lefevre, and Olivier Pietquin. User simulation in dialogue systems using inverse reinforcement learning. In *Twelfth Annual Conference of the International Speech Communication Association*, 2011.

Xiuli Chen, Gilles Bailly, Duncan P Brumby, Antti Oulasvirta, and Andrew Howes. The emergence of interactive behavior: A model of rational menu search. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, pages 4217–4226, 2015.

Pedram Daee, Tomi Peltola, Aki Vehtari, and Samuel Kaski. User modelling for avoiding overfitting in interactive knowledge elicitation for prediction. In *23rd International Conference on Intelligent User Interfaces*, pages 305–310, 2018.

Allan Dafoe, Yoram Bachrach, Gillian Hadfield, Eric Horvitz, Kate Larson, and Thore Graepel. Cooperative ai: machines must learn to find common ground, 2021.

Finale Doshi-Velez and George Konidaris. Hidden parameter markov decision processes: A semiparametric regression approach for discovering latent task parametrizations. In *IJCAI: proceedings of the conference*, volume 2016, page 1432. NIH Public Access, 2016.

Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *International conference on machine learning*, pages 1126–1135. PMLR, 2017.

Alexandre Galashov, Jonathan Schwarz, Hyunjik Kim, Marta Garnelo, David Saxton, Pushmeet Kohli, SM Eslami, and Yee Whye Teh. Meta-learning surrogate models for sequential decision making. *arXiv preprint arXiv:1903.11907*, 2019.

Marta Garnelo, Dan Rosenbaum, Christopher Maddison, Tiago Ramalho, David Saxton, Murray Shanahan, Yee Whye Teh, Danilo Rezende, and SM Ali Eslami. Conditional neural processes. In *International Conference on Machine Learning*, pages 1704–1713. PMLR, 2018a.

Marta Garnelo, Jonathan Schwarz, Dan Rosenbaum, Fabio Viola, Danilo J Rezende, SM Eslami, and Yee Whye Teh. Neural processes. *arXiv preprint arXiv:1807.01622*, 2018b.

Samuel J Gershman, Eric J Horvitz, and Joshua B Tenenbaum. Computational rationality: A converging paradigm for intelligence in brains, minds, and machines. *Science*, 349(6245):273–278, 2015.

Erin Grant, Chelsea Finn, Sergey Levine, Trevor Darrell, and Thomas Griffiths. Recasting gradient-based meta-learning as hierarchical bayes. *arXiv preprint arXiv:1801.08930*, 2018.

Michael U Gutmann, Jukka Corander, et al. Bayesian optimization for likelihood-free inference of simulator-based statistical models. *Journal of Machine Learning Research*, 2016.

Eric J Horvitz, John S Breese, David Heckerman, David Hovel, and Koos Rommelse. The Lumiere project: Bayesian user modeling for inferring the goals and needs of software users. *arXiv preprint arXiv:1301.7385*, 2013.

Ahmed Hussein, Mohamed Medhat Gaber, Eyad Elyan, and Chrisina Jayne. Imitation learning: A survey of learning methods. *ACM Computing Surveys (CSUR)*, 50(2):1–35, 2017.

Antti Kangasrääsiö, Jussi PP Jokinen, Antti Oulasvirta, Andrew Howes, and Samuel Kaski. Parameter inference for computational cognitive models with approximate Bayesian computation. *Cognitive science*, 43(6):e12738, 2019.

Hyunjik Kim, Andriy Mnih, Jonathan Schwarz, Marta Garnelo, Ali Eslami, Dan Rosenbaum, Oriol Vinyals, and Yee Whye Teh. Attentive neural processes. *arXiv preprint arXiv:1901.05761*, 2019.

Richard L Lewis, Andrew Howes, and Satinder Singh. Computational rationality: Linking mechanism and behavior through bounded utility maximization. *Topics in cognitive science*, 6(2):279–311, 2014.

Sheng Li and Handong Zhao. A survey on representation learning for user modeling. In *IJCAI*, pages 4997–5003, 2020.

Hee-Seung Moon, Seungwon Do, Wonjae Kim, Jiwon Seo, Minsuk Chang, and Byungjoo Lee. Speeding up inference with user simulators throughpolicy modulation. In *CHI Conference on Human Factors in Computing Systems*, pages 1–21, 2022.

Andrew Y Ng, Stuart J Russell, et al. Algorithms for inverse reinforcement learning. In *Icml*, volume 1, page 2, 2000.

Alex Nichol and John Schulman. Reptile: a scalable metalearning algorithm. *arXiv preprint arXiv:1803.02999*, 2 (3):4, 2018.

Tomi Peltola, Mustafa Mert Çelikok, Pedram Daee, and Samuel Kaski. Interactive ai with a theory of mind. In *Computational Modeling in Human-Computer Interaction*, 2019.

David Premack and Guy Woodruff. Does the chimpanzee have a theory of mind? *Behavioral and brain sciences*, 1 (4):515–526, 1978.

Neil Rabinowitz, Frank Perbet, Francis Song, Chiyuan Zhang, SM Ali Eslami, and Matthew Botvinick. Machine theory of mind. In *International conference on machine learning*, pages 4218–4227. PMLR, 2018.

Iris Rubi Seaman, Jan-Willem van de Meent, and David Wingate. Nested reasoning about autonomous agents using probabilistic programs. *arXiv preprint arXiv:1812.01569*, 2018.

Scott A Sisson, Yanan Fan, and Mark Beaumont. *Handbook of approximate Bayesian computation*. CRC Press, 2018.

DJ Strouse, Kevin McKee, Matt Botvinick, Edward Hughes, and Richard Everett. Collaborating with humans without human data. *Advances in Neural Information Processing Systems*, 34:14502–14515, 2021.

Mikael Sunnåker, Alberto Giovanni Busetto, Elina Numminen, Jukka Corander, Matthieu Foll, and Christophe Dessimoz. Approximate bayesian computation. *PLoS computational biology*, 9(1):e1002803, 2013.

Brandon M Turner and Trisha Van Zandt. Hierarchical approximate bayesian computation. *Psychometrika*, 79 (2):185–209, 2014.

Jiayu Yao, Taylor Killian, George Konidaris, and Finale Doshi-Velez. Direct policy transfer via hidden parameter markov decision processes. In *LLARLA Workshop, FAIM*, volume 2018, 2018.

Zeping Yu, Jianxun Lian, Ahmad Mahmoody, Gongshen Liu, and Xing Xie. Adaptive user modeling with long and short-term preferences for personalized recommendation. In *IJCAI*, pages 4213–4219, 2019.

Fajie Yuan, Xiangnan He, Alexandros Karatzoglou, and Liguang Zhang. Parameter-efficient transfer from sequential behaviors for user modeling and recommendation. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1469–1478, 2020.

Tan Zhi-Xuan, Jordyn Mann, Tom Silver, Josh Tenenbaum, and Vikash Mansinghka. Online bayesian goal inference for boundedly rational planning agents. *Advances in Neural Information Processing Systems*, 33:19238–19250, 2020.