Learning Human Preferences for Personalized Assistance in Household Tasks

Daphne Chen, Michelle Zhao, Reid Simmons

Carnegie Mellon University,
The Robotics Institute,
Pittsburgh, Pennsylvania, USA
daphnechen@cmu.edu, mzhao2@cmu.edu, reids@cs.cmu.edu

Abstract

As assistive agents become increasingly ubiquitous in home environments, it is ever more important that they are designed to operate within the preferences of the humans around them. Furthermore, these methods should present minimal burden to the user, thus it is key to be able to learn in a data-efficient manner. In this work we propose a study design to learn and evaluate a few-shot model of personalized task preferences within a temporal context that transfers across novel household environments. We propose a method for generating a customizable quantity of synthetic data that reflects the variability in task execution styles seen in the real-world task, and enables us to train a baseline sequential model that predicts the next action a participant will take within a cooking activity. Finally, we present a user study design for evaluating this method with human participants to determine whether the personalized model provides guidance that is preferable over the baseline.

Introduction

In order for assistive agents to effectively operate with humans, it is important that they are capable of learning in a manner that can both rapidly adapt and align with the preferences of unique individuals. Furthermore, to be practical in the real world, assistive agents should be able to learn from few initial observations so as minimize burden on the user. Thus we first describe a technique to learn preferences without labeled human data, and subsequently propose a method that learns preferences in a data-efficient manner using Few-Shot Learning. Lastly, we outline a user study design and empirical evaluation on a household task, and suggest future directions for this work.

Understanding individual preferences is key to providing assistance to a diverse population of users, especially in highly sensitive applications such as caregiving. For example, consider a scenario in which a robot must learn when to intervene if a user performs an unexpected action within a daily routine. In such a setting, it is not always appropriate to assume that the human performs optimally (Carroll et al. 2019). However, learning a model for assistance that incorporates the user's unique preferences and abilities remains challenging. We focus on household tasks in order to constrain the definition of task preferences and ground this work in a natural setting for user assistance.

Household tasks within everyday routines often follow a deterministic structure, or order, from which few deviations are allowable in order to achieve the goal of the task. However, there is enough variation within the task structure to allow for individual preferences to become evident while still being acceptable. We acknowledge that the notion of human preferences is both task-dependent and broad, and may include personality-based, experience-based, and environment-based constraints, among others.

In this study, we specify our definition of *preferences* as temporal patterns of behavior, i.e. the predominantly-chosen order of the user's actions while performing a step-by-step task. The primary research question that this aims to address is *Can an individual's temporal preferences for a household task be learned from limited observations, such that the suggestions from a personalized sequential prediction model are preferable over the baseline?*

Thus in this work we propose a study design that addresses the above by learning user preferences within a cooking task, where the actions, objects, and ingredients are chosen from one set and the task has a fixed outcome. Our long-term goal in for this approach is to build a method that quickly learns user preferences across different environments and evaluate it in a large-scale, real-world study among a diverse population of users.

Related Work

Our work builds upon prior studies in few-shot learning, sequential modeling, and preference learning. This section provides an overview of these areas in relation to this study.

Few-Shot Learning

Few-Shot Learning (FSL) is a paradigm where a pre-trained model sees a smaller set of examples (called a support set) to generalize to a new, similar kind of prediction task on novel data (Vinyals et al. 2016). It is a subset of transfer learning, where a generalized model is further fine-tuned to adapt to novel yet in-distribution tasks. These methods are becoming more widely used because of their ability to learn in a sample-efficient manner.

Sequential Task Prediction

In (Ravichandar et al. 2016) a Long Short Term Memory (LSTM) network is the basis for learning the underlying se-

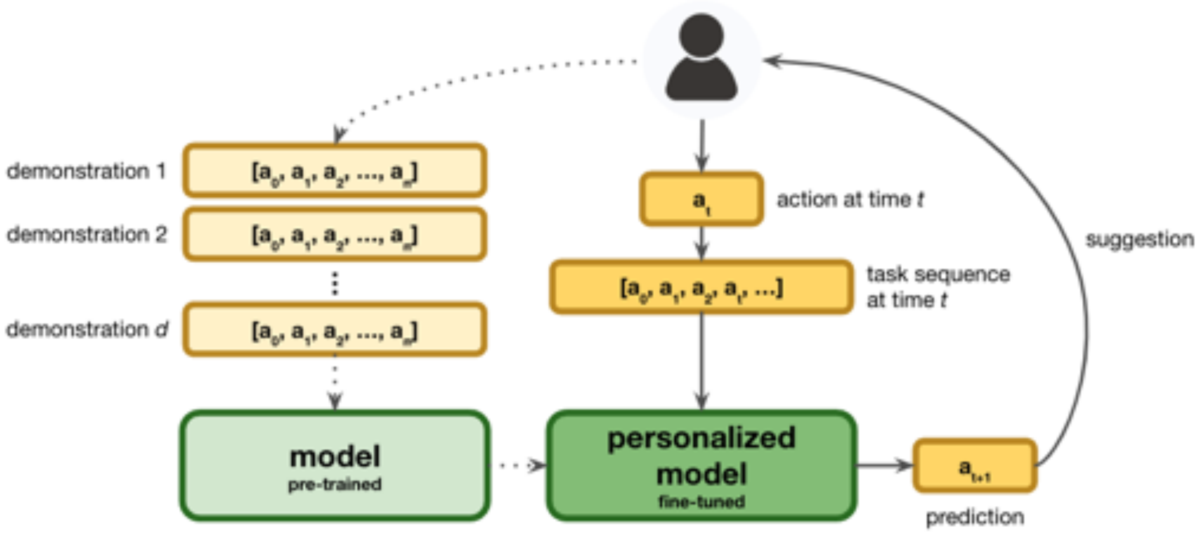

Figure 1: Overview of the study design. Each participant will provide a set of demonstrations offline (left, dotted line), which will be used to fine-tune a pre-trained model. The resulting personalized model incorporates the participant's temporal preferences for executing the task. As the participant performs the task, the personalized model provides the user with personalized suggestions for what action to perform next (right, solid line) based on the predicted action at the subsequent timestep.

quence of steps to predict a human's future actions within a task. While this work addresses sequential action-based prediction, it relies on a deep network to effectively learn both the goal of the task and the remainder of the sequence. In comparison, we assume the task is fixed thus the goal is known, and approach the problem from the perspective of learning within a low-data regime.

Adaptive Assistance

Prior studies on adaptation for assistive tasks have primarily focused on physical assistance, where preferences are explicitly parametrized (Pignat and Calinon 2017). In contrast, this work takes advantage of the ability of Hidden Markov Models to learn *latent* patterns, which may be representative of abstract preferences within sequential behaviors. Models can also be conditioned on latent actions learned from visual inputs in order to generalize to new objects when providing assistance (Karamcheti et al. 2021). We focus our study design towards adapting to latent preferences such that the method can generalize across different people within a task.

Learning Human Preferences

Current methods of learning human preferences often require frequent queries in order to reliably adapt, posing a significant burden to the user. For example, in (Christiano et al. 2017) the authors demonstrate that it is possible to learn from human feedback in a sample-efficient manner – defined as < 1000 bits of feedback – but their method requires continual querying of the human in order to refine the policy.

Recent work (Ouyang et al. 2022) has shown that it is possible to fine-tune large pre-trained language models to generalize to preferences across different user groups, where the output generated by the fine-tuned model is rated more highly than that of baseline language models. However, this approach relies on collecting a large labeled dataset of optimal demonstrations, which is often not feasible in practice.

(Hejna III and Sadigh 2022) treat few-shot preference learning as a multi-task learning problem in order to bypass the need to minimize queries. Instead, they take advantage of the paradigm of shared structure within real-world tasks to meta-learn reward functions from multi-task data. Under the assumption that tasks remain within this expected distribution, they show that their method can rapidly adapt to new preferences from few queries. However, this work defines a preference as a pairwise comparison between two trajectories for short time horizon tasks, making it difficult to determine how the method would work in more complex, longitudinal tasks without continuous feedback.

In contrast, we aim to learn user preferences for sequential tasks over a longer time horizon. We plan to use advances in few-shot learning to adapt to the user in a sample-efficient manner, and apply this method across varied household environments to evaluate its ability to generalize.

Methods

Dataset and Preprocessing

Our criteria for determining a dataset was that it must contain several instances of one task, rather than a few instances of many tasks, in order to place the emphasis on learning user preferences within a single activity. We specifically looked for datasets that encompass cooking tasks because recipes typically follow a predictable structure in their number and order of steps. Despite this, many cooking tasks can also be completed with sufficient variation such that distinctive temporal preferences may arise without diverging from the end goal. For example, within a recipe for preparing vegetable stew, an individual may prefer to peel all vegetables, then cut all vegetables, and finally add them to the pot together; others may prefer to process each ingredient individually, i.e. first peel/cut/add the carrots, then peel/cut/add the potatoes, and lastly peel/cut/add the onions.

For these reasons, we chose the 50 Salads Dataset (Stein and McKenna 2013) as the basis for this study's task. The 50 Salads Dataset contains RGB-D videos of 25 people performing 2 instances of the same salad-preparation task, yielding a total of 50 videos. The videos include timestamped annotations, accelerometer data, and depth maps. To create the initial dataset for this study, we extracted the raw text annotations for each video and used these to represent each instance of the task.

A limitation of this dataset is its small size relative to datasets typically used for training deep models. This problem is not unique to this domain, as labeled real-world data is often difficult and expensive to acquire. Thus in the following section we outline our method for addressing this by developing a generative model of activity sequences to yield a larger, synthetic dataset for the same task.

Dataset Augmentation

In order to closely reproduce the natural variation seen within the original dataset and in the real-world population, we approach the synthetic data generation problem through the lens of a Markov process. We first create a probability matrix based on the transition counts between each possible pair of annotations, or actions, in the sequences of the original dataset. The transition probability at the (i, j)-th index represents the probability of performing action j following action i. These counts are then normalized so that the corresponding probabilities for each annotation sum to 1. Using the action transition matrix, sequences are probabilistically generated via sampling such that they are complete (i.e., achieve the task) and valid. To enforce the validity of each generated sequence and avoid extraneous repetition, we use a constraint tree to eliminate implausible transitions (e.g. mixing an ingredient before it has been added to the bowl). Our task-based constraint tree allows for incomplete or invalid sequences to be pruned, while including probabilistically unlikely sequences to account for a broad range of preference types in the synthetic dataset.

Using this method, we create an augmented dataset containing synthetic sequences for the salad preparation task, which serves as our augmented training set for the few-shot learning model. The next section details the analysis of the dataset to identify patterns of preferences within this task.

Identifying Preference Types

Hidden Markov Models (HMMs) are a statistical method to represent sequential processes as outputs generated by unobservable, internal states (Baum and Petrie 1966). Training an HMM on the augmented dataset of cooking demonstrations seeks to identify underlying temporal processes in the activity, related to key components of the task that are not captured by lower-level actions. We trained an HMM using the `hmmlearn` library, and subsequently generated hidden state sequences over the synthetic data.

Using the strategy identification approach described in (Zhao, Simmons, and Admoni 2022), we train an HMM on the augmented dataset of demonstrations, which are action sequences achieving the cooking task. We obtain a low-dimensional representation of the demonstrations by computing the Viterbi sequences (Forney 1973). Subsequently, we determined latent strategy groups, or preference groups, by applying K-Means clustering to these low-dimensional hidden state sequences. We used the elbow method with silhouette score as the metric to determine the optimal number of clusters $K = 3$. After applying K-Means to the data, we analyzed the sequences within each resulting cluster to determine latent patterns that could be representative of discrete preference groups. Each cluster corresponds to a latent preference type (Figure 2).

As previously described, we define preferences within this study in terms of chosen temporal ordering. Thus we hypothesize that preferences will fall into one of 3 main categories:

P1: The participant makes the salad dressing and subsequently prepares the core ingredients (peel/chop the cucumber/lettuce/tomato) before mixing everything

P2: The participant prepares the core ingredients (peel/chop the cucumber/lettuce/tomato) and subsequently makes the salad dressing before mixing everything

P3: The participant interleaves actions from both salad dressing preparation and core ingredient preparation throughout the task

Modeling Sequential Prediction

The base model used in this study is a Long Short Term Memory (LSTM) network (Hochreiter and Schmidhuber 1997). LSTMs are widely used for sequential modeling because they employ an attention mechanism that learns variable-range long-term dependencies by using previous history to inform the current prediction. Since this study aims to learn the temporal context within a sequence, LSTMs are a viable first candidate towards the problem of predicting the most probable next action within a sequential household task.

The LSTM will serve as the base model for the few-shot learning paradigm. We conducted a preliminary exploration of an LSTM trained on aggregate data compared to individual support sets. For this initial study, we use a two-layer

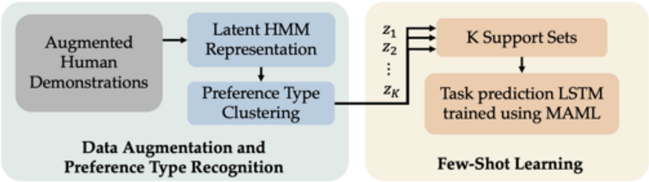
Augmented Human Demonstrations
Latent HMM Representation
Preference Type Clustering
Data Augmentation and Preference Type Recognition
z_1
z_2
⋮
z_K
K Support Sets
Task prediction LSTM trained using MAML
Few-Shot Learning

Figure 2: The preference type identification and few-shot learning model are trained offline.

Table 1: Preliminary Comparison of LSTM Training Paradigms.

Hyperparameters	LSTM-Aggregate	LSTM-Support Sets
Layers	2	2
Hidden Units	128	128
Learning Rate	0.01	0.01
Epochs	1000	1000
Optimizer	Adam	Adam
Acc. Train	0.61	0.29
Acc. Test	0.57	0.21

LSTM with 128 hidden units; hyperparameter details are in Table 1. First, we trained a model, LSTM-Aggregate, on all demonstrations in the synthetic dataset. Using the same network architecture, we trained an ensemble model, LSTM-Support Sets, on $K = 3$ support sets. LSTM-Support Sets is composed of $K = 3$ networks, each one is trained separately on only the data in the corresponding support set. We found that this training paradigm for LSTM-Support Sets led to lower accuracy, likely due to overfitting (Table 1). While we would like for a model to be more granular to different preference types, we find that training separate models does not perform as well as the aggregate model. This indicates that a model that leverages all training data and is optimized for adaptation using a few-shot learning approach, such as Model-Agnostic Meta-Learning (Finn, Abbeel, and Levine 2017), may lead to better predictive accuracy.

Few-Shot Learning

In order to create the support sets, we used our previously-described clustering approach to partition the hidden state sequences into K distinct clusters. This mechanism finds hidden state sequences that are closest in similarity based on the Euclidean distance metric. Our goal for performing clustering on the hidden state sequences rather than the original sequences of data was to determine whether latent patterns would arise from the clusters of hidden states, where each cluster may indicate a unique task execution style.

Each cluster represents a distinct type of preference for performing the salad-making task, which constitute the task set for training an LSTM using a Model-Agnostic Meta-Learning (MAML) (Finn, Abbeel, and Levine 2017) approach. Predicting user actions of a certain preference type is considered an individual *task*. By using MAML for few-shot learning, we aim to train a single model that is sensitive to changes in the input such that few gradient updates can more substantially correct predictions in the direction of the gradient loss. We aim to investigate how a model optimized for rapid adaptation may affect or improve personalized assistance for cooking tasks.

Study Design

Participants

The participants for this study will be recruited from the Carnegie Mellon University student community. We will recruit 20 individuals, taking care to account for a variety of demographics in terms of gender identity, field, and age in order to ensure that user preferences are representative of a diverse population.

Task Definition

The task that participants will be asked to perform during the user study is a short cooking routine comprised of two distinct sub-tasks (e.g. preparing both a salad and a sandwich). We chose the cooking domain because it illustrates common household activities with which most participants will be familiar, while possessing enough variation in steps such that it is possible for unique temporal preferences to emerge.

Procedure

The study will be conducted across two phases. The participant will be given a short tutorial task to familiarize themselves with the environment before beginning the study. In the first phase, participants will be instructed to provide 3 demonstrations of the cooking task in the AI2-THOR household simulator (Kolve et al. 2017). A personalized model M_p will be fine-tuned from the demonstrations. In the second phase, participants will perform the task (1) once without guidance, (2) once while guided by suggestions from the baseline model M_b (Assistant 1), and (3) once while receiving guidance from the personalized model M_p (Assistant 2). The kitchen environments will be randomly chosen from five available floor plans in order to minimize familiarity bias. After interacting with each assistive agent, the participant will be asked to rate their experience by responding to the following questions on a 7-point Likert scale:

1. Assistant $\{1, 2\}$ provided intuitive suggestions.
2. The guidance from Assistant $\{1, 2\}$ was more relevant to me than the guidance from Assistant $\{2, 1\}$.
3. The task was easier when guided by Assistant $\{1, 2\}$.

Experiment Variables

This study will compare assistive agents based on the following two models:

1. *Baseline* – This model M_b will be trained on the aggregate dataset of generated samples, and will thus reflect a generic manner of performing the task.
2. *Personalized* – This model M_p will be fine-tuned on demonstrations from each participant, and will thus incorporate their individual preferences for the task.

Comparing the study participants' experience, both quantitatively and subjectively, will allow us to observe whether a few-shot personalized model successfully provides measurable benefit to the user. The evaluation metrics are outlined in the following section.

Metrics

Quantitative

- *Task efficiency* – measured by the duration of task execution and the number of total actions taken to complete the task.

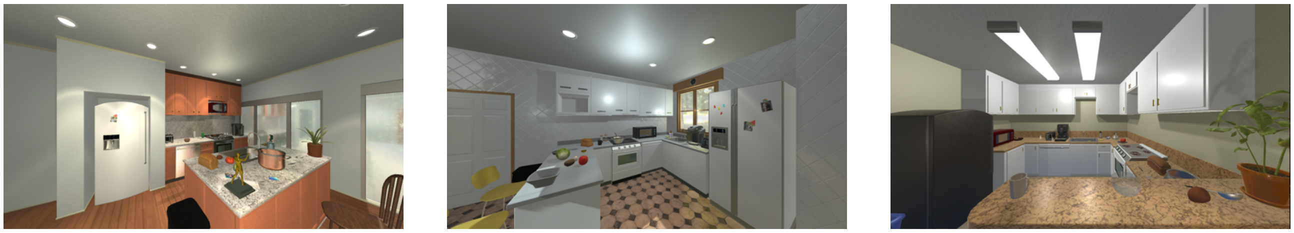

Figure 3: Different floor plans for the kitchen environment in the AI2-THOR simulator.

- *Accuracy* – measured by the number of correct ground truth predictions.

Qualitative

The qualitative metrics will be collected on a 7-point Likert scale and will be based off the responses to the survey questions described in the Procedure section.

- *Intuitiveness* – To what extent are the suggestions provided by M_b and M_p easily understood by the participant?
- *Relevance* – Compared to receiving no guidance, to what extent do the suggestions provided by M_b and M_p help the participant to perform their intended action towards completing the task?
- *Alignment* – To what extent is the guidance from M_p perceived as better than that of M_b; i.e. more aligned with the user's task preferences?

Hypotheses

H1: The participants will prefer the suggestions provided by the personalized model over those from the baseline model.

H2: The participants who are guided by the personalized model will have higher efficacy in performing the task.

H3: The personalized model will be more accurate than the baseline model at predicting the participants' next action.

Proposed Validation and Analysis

We propose to validate the described approach through the ANOVA test to determine the statistical significance of the results against the hypotheses. We will evaluate the quantitative performance of our method over the baseline in terms of both prediction accuracy and participant efficacy. Additionally, we will evaluate the qualitative results by comparing the participant responses to the questions defined above.

Conclusions and Future Work

In this work we present a study design for implementing and validating a method that learns temporal preferences from limited data, and sequentially provides personalized suggestions towards completing the task. We propose an evaluation on a household activity in a simulated home environment, including metrics that will indicate the efficacy and subjective user preference of this approach.

Pending encouraging results from this study, we plan to next investigate more complex sequential models, such as pretrained transformer-based architectures which have indicated promising ability to encapsulate real-world semantic context across a variety of domains. This direction may allow us to address another area of future work, which is to evaluate our approach on long-horizon tasks such as day-long or multi-day household routines.

Finally, we hope that this work will have downstream applications towards supporting individuals in leading independent lives through autonomous, at-home assistance that is personalized to their needs.

Acknowledgments

The authors would like to thank the NSF AI-CARING Institute for supporting this work (Grant No. 53668.1A.1043222). Additionally, the authors are grateful to Gavin Zhu for the assistance in setting up the simulator used in pilot trials of this study.

References

Baum, L. E.; and Petrie, T. 1966. Statistical Inference for Probabilistic Functions of Finite State Markov Chains. *The Annals of Mathematical Statistics*, 37(6): 1554–1563.

Carroll, M.; Shah, R.; Ho, M. K.; Griffiths, T.; Seshia, S.; Abbeel, P.; and Dragan, A. 2019. On the Utility of Learning about Humans for Human-AI Coordination. In Wallach, H.; Larochelle, H.; Beygelzimer, A.; d'Alché-Buc, F.; Fox, E.; and Garnett, R., eds., *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc.

Christiano, P.; Leike, J.; Brown, T. B.; Martic, M.; Legg, S.; and Amodei, D. 2017. Deep reinforcement learning from human preferences.

Finn, C.; Abbeel, P.; and Levine, S. 2017. Model-agnostic meta-learning for fast adaptation of deep networks. In *International conference on machine learning*, 1126–1135. PMLR.

Forney, G. 1973. The viterbi algorithm. *Proceedings of the IEEE*, 61(3): 268–278.

Hejna III, D. J.; and Sadigh, D. 2022. Few-Shot Preference Learning for Human-in-the-Loop RL. In *6th Annual Conference on Robot Learning*.

Hochreiter, S.; and Schmidhuber, J. 1997. Long Short-Term Memory. *Neural Computation*, 9(8): 1735–1780.

Karamcheti, S.; Zhai, A. J.; Losey, D. P.; and Sadigh, D. 2021. Learning visually guided latent actions for assistive teleoperation. In *Learning for Dynamics and Control*, 1230–1241. PMLR.

Kolve, E.; Mottaghi, R.; Han, W.; VanderBilt, E.; Weihs, L.; Herrasti, A.; Gordon, D.; Zhu, Y.; Gupta, A.; and Farhadi, A. 2017. AI2-THOR: An Interactive 3D Environment for Visual AI. *arXiv*.

Ouyang, L.; Wu, J.; Jiang, X.; Almeida, D.; Wainwright, C. L.; Mishkin, P.; Zhang, C.; Agarwal, S.; Slama, K.; Ray, A.; Schulman, J.; Hilton, J.; Kelton, F.; Miller, L.; Simens, M.; Askell, A.; Welinder, P.; Christiano, P.; Leike, J.; and Lowe, R. 2022. Training language models to follow instructions with human feedback.

Pignat, E.; and Calinon, S. 2017. Learning adaptive dressing assistance from human demonstration. *Robotics and Autonomous Systems*, 93: 61–75.

Ravichandar, H. C.; Kumar, A.; Dani, A.; and Pattipati, K. R. 2016. Learning and predicting sequential tasks using recurrent neural networks and multiple model filtering. In *2016 AAAI Fall Symposium Series*.

Stein, S.; and McKenna, S. J. 2013. Combining Embedded Accelerometers with Computer Vision for Recognizing Food Preparation Activities. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '13, 729–738. New York, NY, USA: Association for Computing Machinery. ISBN 9781450317702.

Vinyals, O.; Blundell, C.; Lillicrap, T. P.; Kavukcuoglu, K.; and Wierstra, D. 2016. Matching Networks for One Shot Learning. *CoRR*, abs/1606.04080.

Zhao, M.; Simmons, R.; and Admoni, H. 2022. Coordination with Humans via Strategy Matching.