Improving the Prediction of Individual Engagement in Recommendations using Cognitive Models*

Roderick Seow[1,*], Yunfan Zhao[2], Duncan Wood[1], Milind Tambe[2] and Cleotilde Gonzalez[1]

[1]*Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, Pennsylvania, United States of America*
[2]*Harvard University, Massachusetts Hall, Cambridge, Massachusetts, United States of America*

Abstract

For public health programs with limited resources, the ability to predict how behaviors change over time and in response to interventions is crucial for deciding when and to whom interventions should be allocated. Using data from a real-world maternal health program, we demonstrate how a cognitive model based on Instance-Based Learning (IBL) Theory can augment existing purely computational approaches. Our findings show that, compared to general time-series forecasters (e.g., LSTMs), IBL models, which reflect human decision-making processes, better predict how individuals' behaviors change over time (transition-consistency) and in response to receiving an intervention (intervention-sensitivity). We further show that IBL parameters capture the individual differences in transition-consistency and intervention-sensitivity and that other time series models can use these individual-level IBL parameters to improve their training efficiency.

Keywords

Limited resource allocation, Restless multi-armed bandit, Time-series forecasting, Cognitive modeling, Instance-based learning

CEUR-WS.org/Vol-3823/5_seow_improving_177.pdf

1. Introduction

Public health programs play an essential role in improving the health outcomes of individuals and communities, often through education and subsequent behavioral change. Some health programs interact with their intended beneficiaries in a broad and infrequent manner. For example, a campaign about the health risks of smoking may address a general population of smokers through scattered advertisements in the media [1]. Others rely on repeated direct interactions with their intended beneficiaries—for example, a maternal health program that sends automated messages about exercise and nutrition to enrolled expectant mothers [2]. In this case, it is crucial that mothers remain engaged for the duration of the program or as long as possible to receive the maximum benefit. Unfortunately, many such programs face high levels of disengagement and dropout, which severely limit their effectiveness [3].

To reduce dropout, programs can provide interventions designed to increase beneficiary engagement. For example, program staff can make personalized service calls to beneficiaries at risk of dropping out to address concerns about declining participation. However, programs with limited staff and resources cannot provide interventions to all their beneficiaries. Instead, they must choose a subset of beneficiaries to receive the intervention.

Recent approaches have adopted the Restless Multi-Armed Bandit (RMAB) framework [4]. Classically, the RMAB framework models the engagement dynamics of each beneficiary (i.e., the transition between discrete engagement and disengagement states) as a Markov Decision Process (MDP). A centralized planner then recommends which beneficiaries to intervene on at each time point based on the dynamics learned from the MDPs. Importantly, beneficiaries can change their engagement state even without an intervention. The best-performing approach to this framework is the Whittle index [5], which ranks beneficiaries by their likelihood of becoming and remaining engaged after getting an intervention.

HealthRecSys'24: The 6th Workshop on Health Recommender Systems co-located with ACM RecSys 2024
*Corresponding author.
yseow@andrew.cmu.edu (R. Seow); yunfanzhao@fas.harvard.edu (Y. Zhao); djwood@andrew.cmu.edu (D. Wood); milind_tambe@harvard.edu (M. Tambe); coty@cmu.edu (C. Gonzalez)
© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073

This framework assumes that beneficiaries' dynamics—how their level of engagement changes over time—can be modeled as a MDP. However, recent work has shown that these dynamics are often non-Markovian; that is, a beneficiary's transition between being engaged and disengaged (their state) depends on their individual history [6]. As a substitute, Danassis and colleagues (2023) developed the Time-series Arm Ranking Index (TARI), which allows non-Markovian dynamics and continuous engagement levels within the RMAB framework. The TARI approach models beneficiaries using time-series forecasters such as LSTMs [7] or transformers [8] instead of MDPs. Using these learned time-series models, TARI computes an index that captures the relative engagement benefit of intervening on a beneficiary by comparing their expected future engagement with and without an intervention at the current time point and then ranks beneficiaries by the computed indices.

Within the TARI approach to the RMAB framework, we propose modeling individual beneficiaries with computational cognitive models instead of LSTMs. Specifically, we use models based on Instance-Based Learning Theory (IBLT) [9] to represent each beneficiary's time series activities. We expect using an IBL model will convey two advantages over general time-series forecasters (TSFs).

First, IBL models are personalized to the individual instead of trained across the entire dataset, allowing the models to capture individual differences in behavioral dynamics (i.e., individual-level differences in the response to interventions and general engagement patterns). Because TSFs need large amounts of data to fine-tune their many parameters and there is relatively sparse data available per beneficiary, existing approaches train a single model using the combined data from all available beneficiaries. For instance, Danassis and colleagues (2023) create a set of fixed-length vectors from each beneficiary's trajectory. Each vector represents the intervention history and engagement levels of seven consecutive timesteps. Then an LSTM trains on these vectors to predict the engagement level at the next timestep given the previous seven [6]. In contrast, an IBL model has enough cognitively grounded structure that it can be applied to an individual with sparse data [10]. This is because IBL theory makes strong assumptions about the cognition of human decision making derived from prior research. In particular, IBL theory accounts for memory effects (i.e., the more recent a data point is, the more it contributes to the prediction) and similarity effects (i.e., engagement levels that are observed under similar contexts should be similar to each other) that general TSFs do not.

Second, the parameters of IBL cognitive models are grounded in psychological constructs such as memory and attention, making them more interpretable than the parameters of general LSTMs. This advantage allows cognitive models to shed light on the individual differences that drive beneficiaries' behaviors. We demonstrate how these individual differences can be used to augment existing approaches that rely primarily on general LSTMs.

1.1. Our Contributions

1. Using real-world data from a maternal health program, we show that cognitive models more accurately capture the temporal dynamics of individual behaviors and predict the behaviors of individual beneficiaries than LSTMs.
2. We show that personalized cognitive models reveal individual characteristics through weights fitting. We illustrate how to cluster individuals using learned characteristics.
3. We show clustering based on learned individual characteristics could guide and improve the performance of time series models. Specifically, we show that LSTMs trained in clusters better predict behaviors within those clusters compared to LSTMs trained on random samples of beneficiaries.

1.2. Background

1.2.1. Time-series Arm Ranking Index (TARI)

In the context of multiple potential beneficiaries treated as arms in an RMAB, each arm is considered an independent time series [6]. We train a model to predict the next state s_{t+1} based on three pieces of information:

- A historical record (length h) of the past states and actions of the arm (denoted by $\{(s_i, a_i)\}_{i<t}^{h}$).
- The current state (e.g., engagement level) of the arm s_t.
- A potential action a_t, where 1 represents an intervention and 0 represents no intervention.

This training takes place offline. During test time, the model's parameters are fixed, and we use the trained model with iterated multi-step forecasting to generate a long-term forecast of future states $s_{t+1}, s_{t+2}, ..., s_{t+H}$ [11].

For each arm, we use the TSF model to estimate two values:

- Time to disengagement **with** intervention u_n: The predicted number of timesteps until the arm becomes non-engaged if we intervene at the current timestep (t) and never again.
- Time to disengagement **without** intervention v_n: The predicted number of timesteps until the arm becomes non-engaged with no interventions at any point.

The TARI index for an arm is the ratio of these two times. A higher TARI index suggests that intervening now would be more beneficial. Finally, similar to the Whittle index, we choose to act on the k arm with the highest TARI index at each timestep:

$$\text{TARI}(n) = \frac{u_n}{v_n}.$$

1.2.2. Instance-Based Learning Theory

Instance-Based Learning Theory (IBLT) is a cognitive theory of dynamic decision making, grounded in human learning and memory mechanisms. It has successfully accounted for human behavior in a variety of contexts, ranging from abstract repeated-choice tasks to more real-world search and choice [12] and continuous control tasks [9]. In most of these contexts, human participants are required to make a series of decisions while exposed to changing environments.

The essence of IBLT is storing past decisions in the form of an instance. An instance has three parts: the context, the choice made, and the associated utility. For example, these would correspond to the cloudiness of the sky, the decision to bring an umbrella, and the inconvenience of bringing an umbrella when it does not rain (supposing it did not rain in this instance).

IBL agents compute the expected utility of a choice by considering past instances where they made that choice, weighted by the contextual similarity to the present. For example, an IBL agent would compute the expected utility of bringing an umbrella by considering its past experiences with an umbrella, most strongly considering the experiences with the most similar cloudiness to the current moment.

More precisely, the expected value of a choice is the weighted average of the utilities from all the past instances where that choice was made weighted by the retrieval probability (i.e., the probability of remembering that instance) to generate estimates of the choice's expected utility called a blended value. The agent will deterministically make the available choice with the highest computed blended value (expected utility). The utilities of instances in memory are the actual outcome the agent observed when it chooses, regardless of what it predicted. After choosing, a new instance is added to the agent's memory, reflecting what happened. If the exact combination of context, choice, and utility has appeared before, that instance instead has its frequency and recency information updated.

The contribution of an instance's utility to a choice's expected value depends on the instance's memory activation (the salience of that instance given the current context). The activation of an instance (A_i) reflects how readily it comes to mind, which in turn depends on the frequency, recency, and similarity factors according to the following equation as proposed by the ACT-R cognitive architecture [13].

$$A_i = ln(\sum_j (t - t'_{i,j})^{-d}) + \mu \sum_k (w_k(Sim(s_{i,k}, s_{t,k}) - 1) + \sigma\xi \tag{1}$$

Three expressions additively determine an instance's activation (salience in memory). The first expression is the contribution of frequency and recency where t is the current timestep, $t_{i,j}$ is the timestep of the jth appearance of instance i, and d is a decay parameter. The second expression is the contribution of similarity where μ is a scaling parameter on the overall impact of similarity to activation, w_k is the importance (or weight) of an attribute in the similarity calculations, Sim is a similarity function that returns the degree of similarity for a particular attribute k between instance i ($s_{i,k}$) and the current context ($s_{t,k}$). The third expression adds some randomness to the memory retrieval process where $\sigma\xi$ is a scaled noise distribution. The activation of an instance is then converted into a retrieval probability using the Boltzmann softmax function, where τ is the temperature parameter:

$$P_i = \frac{e^{A_i/\tau}}{\sum_j e^{A_j/\tau}} \tag{2}$$

The IBL model computes the expected utility of a choice (a) from the utility u_i of each retrieved instance i weighted by their respective retrieval probability P_i [9, 14]:

$$V(a) = \sum_i P_i u_i \tag{3}$$

The agent then makes the choice with the highest expected utility, $V(a)$.

2. Proposed Approach

We model each beneficiary with their own individual-level IBL model. We do not have the IBL model make a choice. Instead, we use the IBL's expected utility calculation to predict the engagement level in the next period (see Figure 1a). Because we only use the model's blended value, the choice part of the instance is always held constant. The utility of an instance is a beneficiary's engagement level (thus, the expected utilities are weighted averages of past engagement levels then used to predict future engagement levels). The context has two components derived from a beneficiary's history: (1) the engagement level of the previous period and (2) the number of timesteps since the last intervention. The IBL model determines the similarity between the current context and each instance in memory according to these two attributes, and these similarities are factored into an instance's activation according to Equation 1.

The IBL models are trained through model tracing: recording each timestep of an individual beneficiary as an instance with the relevant context (previous engagement level and time since the last intervention) and utility (engagement level). Thus, the model "traces" the beneficiary it is modeling by giving itself memories as if it were the beneficiary that produced the data.

As shown in Equation 1, the similarity of two instances depends not only on the similarity between attribute values, but also on the weight or influence of the attribute (w in Equation 1). Psychologically, an attribute weight represents how relevant the agent considers an attribute to make similarity judgments between two contexts. To capture potential individual differences between beneficiaries, we personalized an IBL model for a specific beneficiary by finding the combination of attribute weight values that resulted in the smallest training prediction error.

Given the predictions of engagement level, we follow TARI and generate a ranked list of beneficiaries according to the predicted relative benefit in the improvement in engagement each beneficiary receives from an intervention versus not receiving one.

3. Simulation Setup

3.1. Data

We tested our approach on data from a maternal healthcare program operated by an NGO called ARMMAN [15]. This program sends automated messages about maternal and infant care to expectant and recent mothers in vulnerable communities. However, over the many weeks of pregnancy and postnatal care, many mothers stop engaging with the automated messages. ARMMAN has a limited capacity to intervene: Program staff can call mothers individually, allowing them to hear from a real person and ask questions, hopefully increasing their future engagement.

ARMMAN collected the data in 2022 from 12,000 mothers over 40 weeks. The data have two recorded values per mother per week: the amount of time she spent listening to the automated health message (engagement level, recorded as a number in $[0, 1]$) and if she received a call from ARMMAN staff (if she received an intervention, recorded as a binary value). The data also contain mothers' demographic information; however, previous work found little or no prediction benefit in incorporating demographic variables as additional features for time-series forecasters [6]. Thus, we do not consider them in this approach.

For an IBL model to predict an intervention's effect on a mother's engagement level, it needs to be trained on a trajectory that includes at least one intervention. However, of the 12,000 mothers, only 5,400 mothers received at least one intervention early on in their program tenure. We also assume that mothers receive an intervention (or the equivalent of an intervention) upon their enrollment in the program. Because the time needed to train all the IBL models scales with the number of mothers, we explored our approach with a subset of 210 mothers out of the 5,400 mothers who received at least one intervention in their actual trajectories.

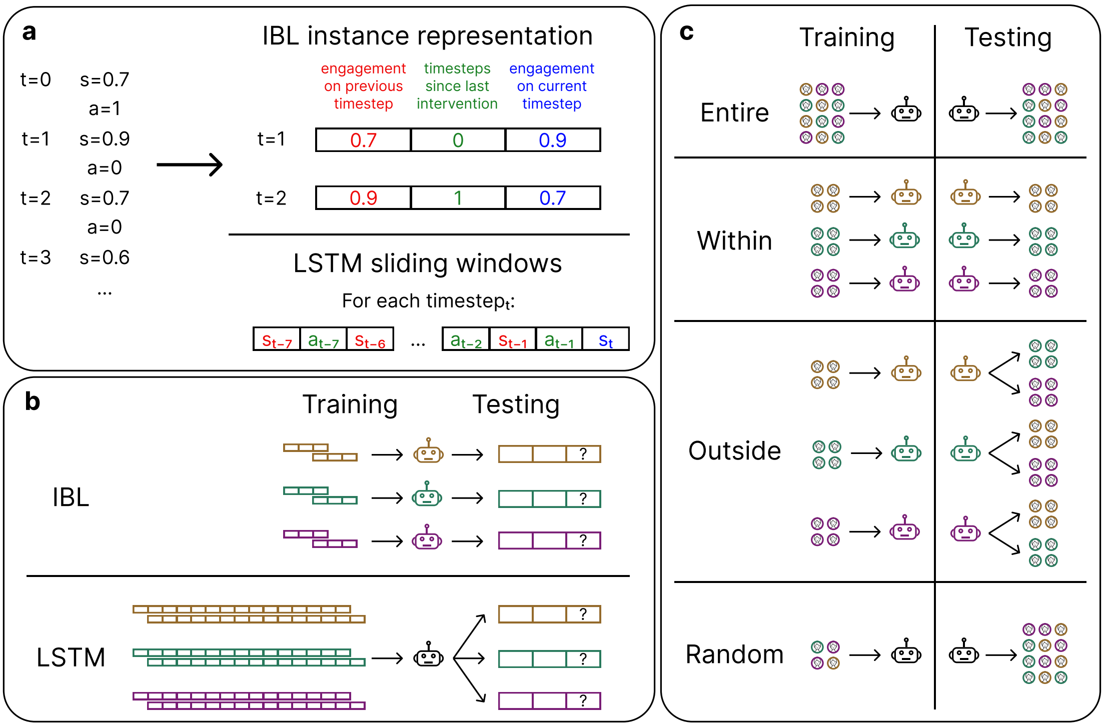

Figure 1: a) Transforming a beneficiary's trajectory into IBL instances and LSTM sliding windows where "s" is state (engagement level) and "a" is action (intervention status). Each IBL instance represents the engagement context, consisting of three parts: the engagement level in the immediately prior timestep, the number of timesteps since the last intervention, and the engagement level in the current timestep. The LSTM receives the state and action of the previous seven timesteps. **b)** Training / testing setup for next-step prediction. Different colors indicate data from individual beneficiaries. One IBL model is trained per beneficiary and the same model predicts for that beneficiary in testing. One LSTM is trained using data from all beneficiaries and that model predicts for all beneficiaries in testing. **c)** LSTM training / testing setups under various IBL-clusters-related conditions. Different colors indicate different clusters. The Random condition LSTM is trained on a random third of the training data to make it comparable to the Within and Outside methods.

3.2. Model Training

We designated the first 25 weeks of each mother's trajectory as training data. Following the approach of previous studies, we reconstructed the training data into sliding windows of 7 consecutive timesteps to train the LSTM model. Then, we trained an LSTM model on the entire training dataset (see Figure 1b for a visual description of the differences between the IBL and LSTM training / testing procedure).

Each IBL model (one for each mother) traced a mother's trajectory through each timestep up to week 25. To find the best-fitting profile of attribute weights per mother, a grid search looked for the combination that minimized the weighted sum of next-step prediction errors (loss). The grid iterates all possible combinations of parameter values in the range (0, 0) to (5, 5) in intervals of 0.5.

An IBL model should improve its predictive accuracy as it accumulates more instances and becomes more "familiar" with the individual it's modeling. So, we weight earlier prediction errors (e.g., at weeks 1 and 2) less than later ones (e.g., at weeks 24 and 25) according to the following equations:

$$loss = \sum_t q_t * SE_t \tag{4}$$

$$q_t = \frac{e^{t/10}}{\sum_i e^{t_i/10}} \tag{5}$$

where SE_t is the squared prediction error and q_t is the normalized weight factor for the timestep t.

3.3. Next-step prediction task

For the next-step prediction task, the IBL and LSTM models generated predictions for the next 14 weeks, starting from week 26. The models iteratively generated predictions for the next timestep (i.e., for week 26, then for week 27, then for week 28, etc.). In our data, ARMANN did not provide interventions to mothers past 14 weeks, so there are no interventions during the testing period for the next-step prediction task. This allows us to compare the models' engagement predictions with the ground truth data without the need to generate synthetic counterfactuals.

3.4. Model comparison with simulated counterfactuals

In a separate analysis, we compared the IBL-TARI and LSTM-TARI policies with three other baseline policies: (i) myopic uniformly random allocation (ii) round-robin allocation, which iteratively cycles through the population, and (iii) no interventions. Consistent with ARMMAN's limited resources, all policies are allowed a 3% budget within a timestep (except the no intervention policy), i.e., 6 beneficiaries per timestep.

To simulate the effect of these policies', we followed previous studies and trained a separate LSTM model on trajectories from the entire dataset (approximately 12000 beneficiaries)—far larger than the policy LSTM's training set—to generate mothers' counterfactual engagement. This counterfactual generator only operates when a mother's simulated trajectory deviates from her ground-truth trajectory (i.e., received a simulated intervention when she did not receive one in reality or not receiving a simulated intervention when she did receive one in reality).

4. Results

4.1. Predicting next-step engagement levels

Figure 2 shows the prediction error of the two types of models during the 14 week the testing period of the next-step prediction task. The IBL models consistently achieve lower prediction errors than the LSTM—about 10% less mean error (0.23 on average) compared to the LSTM model (0.32 on average).

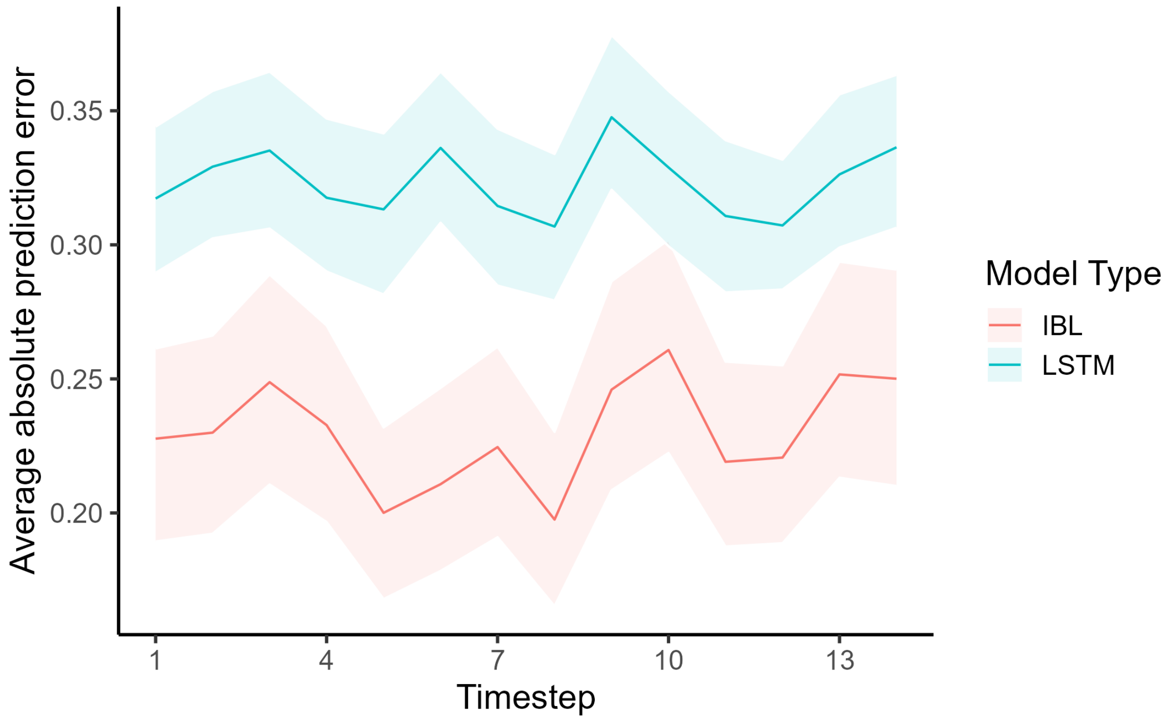

Figure 2: Next-step prediction error per testing timestep (smaller is better). $N = 210$. The personalized IBL models consistently achieve lower average prediction error (0.23 across all timesteps) than the LSTM model (0.32)—a 29% reduction in error.

4.2. Using predictions to inform intervention allocation

Having demonstrated that IBL models yield more accurate next-step predictions of engagement levels compared to the LSTM model, we employed simulations to examine if integrating IBL models with the TARI index would result in higher overall levels of engagement. Figure 7 (in Appendix) shows the percentage of mothers who have an engagement level of 0.25 and above during each test timestep following the various intervention allocation policies. The LSTM-TARI policy outperforms all other policies, particularly later on in the testing period. The average percentage of engaged mothers across testing timesteps for the various policies are as follows: IBL-TARI (55.58%); Round-robin (55.65%); Random (54.29%); LSTM-TARI (61.05%); Control (46.22%). Generating counterfactuals using an LSTM model may explain why the IBL's advantage in next-step prediction does not translate to higher counterfactual engagement in the simulated task.

4.3. Describing individual differences with IBL parameters

Another advantage of the proposed IBL approach over more general TSFs (as exemplified by the LSTM approach) is identifying individual differences in beneficiaries' behavioral characteristics. Because we model each beneficiary with a separate IBL model, the models' parameter values capture how beneficiaries differ from each other in their engagement and intervention dynamics.

Specifically, the attribute weights of each mother's IBL model trained on data from the first 25 weeks tracks (1) how reliably her engagement changes in response to an intervention (intervention-sensitivity) and (2) how consistently her engagement changes between timesteps (transition-consistency), i.e., how similarity between the engagement levels in different timesteps corresponds to similarity between the engagement levels in their following timesteps.

Figure 3 plots mothers according to their attributes weights. Visually, there appear to be three types of mothers. We confirmed that the optimal number of clusters is between 3 to 4 clusters using multiple measures of cluster validity indices (silhouette: 4, within-sum-of-squares: 4, gap statistic: 3).

For interpretability, subsequent analyses assume three clusters.

Each cluster was labeled according to its defining characteristics. Intervention-sensitive mothers (N = 62) have a high intervention lag weight; i.e., a mother's engagement level t timesteps after her most recent intervention will be similar to other times she was t timesteps after the most recent intervention. Transition-consistent mothers (N = 66) have a high previous engagement weight; i.e., if engagement levels at timesteps $i-1$ and $j-1$ (x_{i-1} and x_{j-1}) are similar to each other, then we should expect engagement levels x_i and x_j to also be similar to each other. State-stable mothers (N = 82) have low values in both dimensions; i.e., these mothers will tend to engage at the level they engage most frequently and most recently.

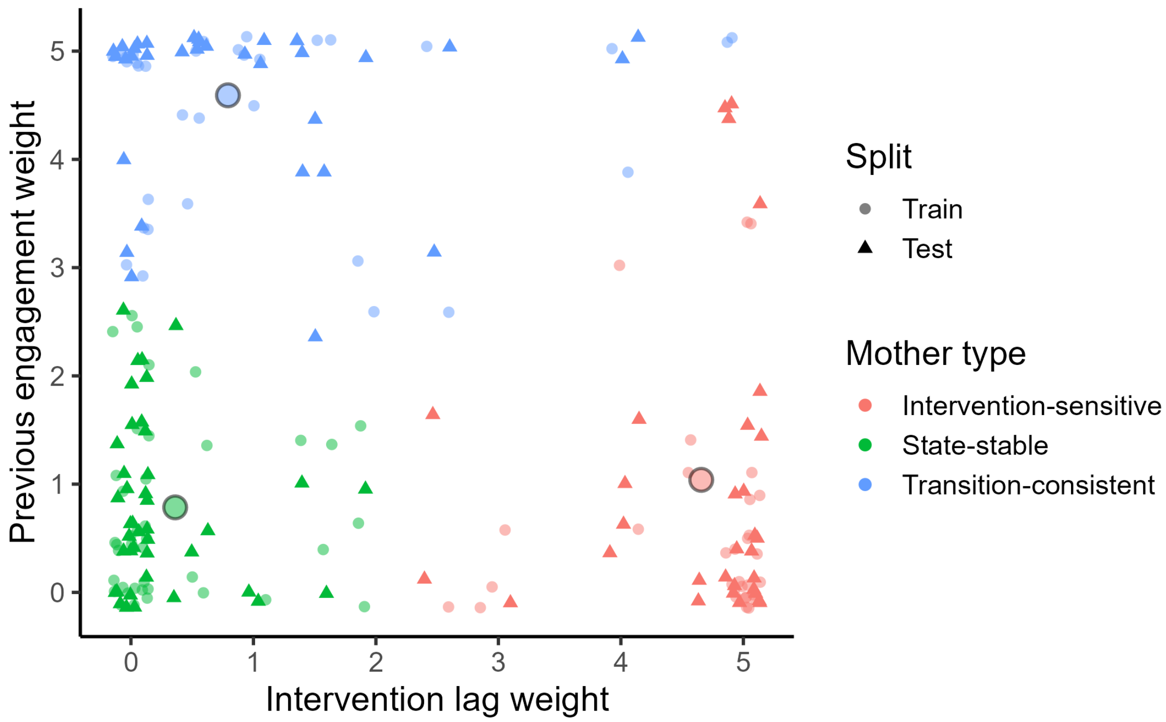

Figure 3: Distribution of attribute weight profiles. Each dot represents a specific mother's IBL model. Points are slightly jittered to better visualize density. Lighter points indicate training set mothers while darker points indicate testing set mothers. Beneficiaries are clustered into 3 clusters according to a k-means algorithm. Larger circles outlined in black indicate cluster centroids.

4.3.1. Testing the generalizability of individual differences in IBL attribute weights

The previous results showed that individual mothers have different optimal weight profiles for the IBL instance similarity attributes. However, we still must show these weights capture individual differences that predict behavioral dynamics beyond an IBL modeling context. If these attribute weights reflect underlying individual differences in engagement and intervention dynamics, then we would expect other modeling approaches that account for these attribute weights to perform better than those that do not.

We tested this hypothesis by training and testing LSTM models with cluster-related samples. Using the cluster assignments described in the previous section, we randomly split each cluster in half to construct a training and a testing set. For all methods, the task is to predict the next-state engagement of the testing set during the testing period (i.e., the final 14 weeks). We examined the following methods (see 1c for a visual description):

- **Entire**: An LSTM trained on all training mothers and predicting for all test mothers.

- **Within-cluster**: An LSTM trained on each cluster and predicting for test mothers in their respective training cluster. We hypothesize that this method should yield the most accurate predictions given the informativeness of the identified IBL clusters.
- **Outside-cluster**: An LSTM trained on each cluster and predicting for test mothers from a different cluster. Because there are 3 clusters, we average prediction error from running the model on both out-of-cluster testing sets.
- **Random**: An LSTM trained on a random subset of one-third of the training set ($N = 35$) and predicting for all test mothers

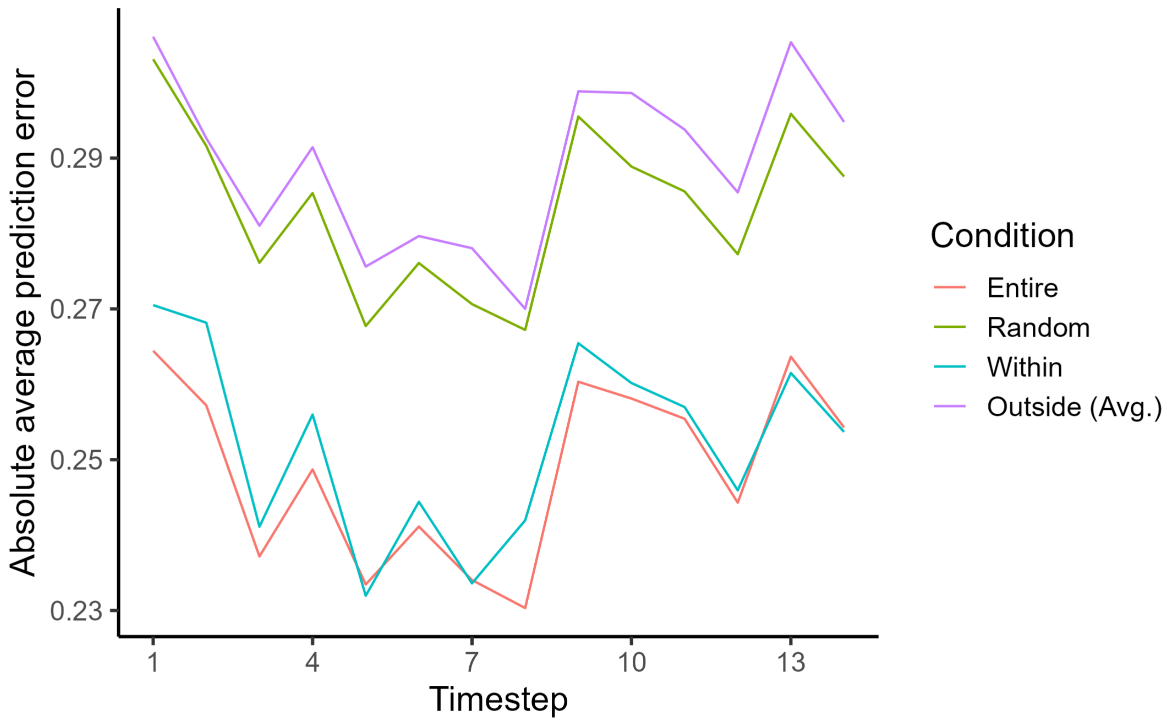

Figure 4: Next-step prediction error per testing timestep for the various LSTM training-testing methods. Outside-cluster method error is the average error of a model trained on each of the three testing sets tested on each of the two other clusters (six total combinations).

Figure 4 shows the average prediction error for the various methods. Supporting the generalizability of the IBL attribute weight clusters, LSTMs trained and tested on different clusters (Outside-cluster) and the LSTM trained on a random sample (Random) performed worse than LSTMs trained and tested on the same clusters (Within).

The Within-cluster and Entire methods yield similar average prediction accuracies (despite each Within-cluster LSTM having less training data). However, as shown in Figure 5, the Within-cluster method produces more accurate predictions than the Entire method for two-thirds of the mothers (Within 70 vs. Entire: 35), which suggests that the clusters effectively categorize the majority of the population.

Follow-up exploratory analyses suggest that mothers closer to their cluster's centroid are better predicted by the Within-cluster method as compared to the Entire method. Using a distinct binomial regression model for each cluster, we used the Euclidean distance between a mother's attribute weights and her centroid to predict the method (Within-cluster or Entire) with lower prediction error. The effects were statistically insignificant, plausibly due to the imbalanced and small classes. Still, for State-stable ($\beta_{dist} = -0.8637$) and Transition-consistent mothers ($\beta_{dist} = -1.235$), the closer they were to the cluster centroid, the more likely the Within-cluster method gave better predictions than the Entire method (see Figure 6).

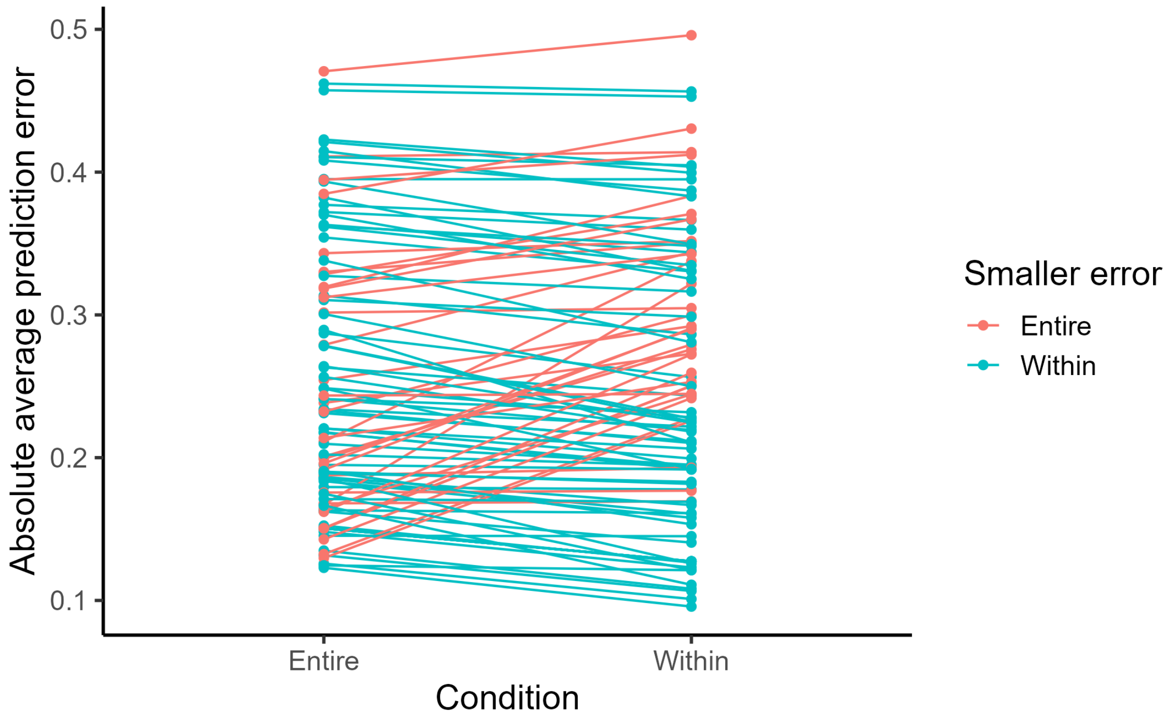

Figure 5: Comparison of prediction errors between Entire and Within-cluster methods. Points show average prediction error for individual mothers in the testing set and lines indicate the same mother across conditions. Blue lines indicate mothers best predicted by the Within-cluster method. Red lines indicate mothers best predicted by the Entire method.

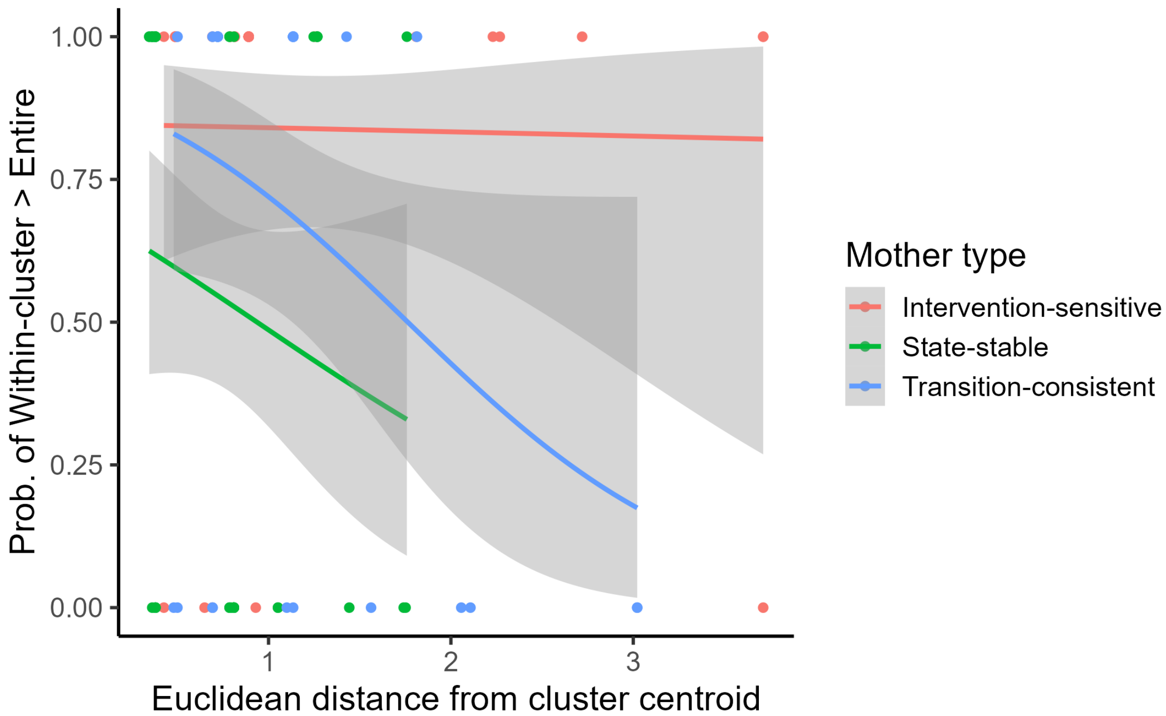

Figure 6: Probability the Within-cluster method had lower prediction error than the Entire method. Probability 1 means the Within-cluster method best predicts a mother's engagement. Lines show predictions by binomial regression models, and shaded areas reflect the 95% CIs.

5. Conclusions and Future work

Public health programs with limited resources to provide interventions must face the difficult problem of determining which beneficiaries to give interventions. Prior work demonstrated that the TARI algorithm—which relies on the predictions of time-series forecasters—outperforms existing recommendation algorithms [6]. To improve the effectiveness of TARI's recommendations through more accurate predictions, we developed a novel approach that uses computational cognitive models derived from Instance-Based Learning Theory [9, 14] to model the human transition dynamics of program engagement. Given the non-Markovian nature of a beneficiary's transitions between states, a cognitive algorithm that reflects the effects of time, attention, and context similarity on human memory and decisions provides more accurate predictions of beneficiaries' engagement dynamics and more meaningfully describes relevant individual differences that other prediction methods can leverage.

We tested our method on real-world engagement data from mothers enrolled in a maternal health program, finding it results in higher next-step prediction accuracy compared to existing time-series methods using LSTM models. The personalized IBL models revealed that mothers clustered into three types based on their IBL attribute weights— specifically, how their engagement patterns could be attributed to intervention-sensitivity and transition-consistency. These clusters generalized to other time-series models: LSTMs trained and tested on within-cluster data outperformed all other LSTM training methods for most mothers.

Future work could explore recommending personalized interventions and intervention schedules based on these identified individual differences in behavioral characteristics. For example, intervention-sensitive beneficiaries could receive more frequent but shorter calls as the model predicts that each call would provide a substantial boost to their engagement levels. In comparison, state-stable beneficiaries who are likely to maintain their engagement levels over time could receive fewer but more intensive calls that maximize their engagement levels by the end of each call. These beneficiaries would then be able to sustain a high engagement level until the next time they receive a call.

One limitation of this work is the counterfactual generator used to simulate the effect of recommended interventions—an LSTM model trained on a much larger dataset. The higher next-step prediction accuracy of the IBL model suggests that the LSTM architecture may systematically fail to pick up certain cognitively-determined regularities in the human data. Thus, the LSTM-generated counterfactuals may also fail to produce simulated data with those regularities. This might explain why IBL's next-step prediction advantage over the LSTM does not translate to better performance in the model comparison with simulated counterfactuals. Future work should prioritize testing our method with other datasets, particularly with a paradigm where the effects of recommended interventions on participants' task engagement can be experimentally measured, rather than relying on plausibly biased counterfactuals.

Another limitation is that our method uses the first 25 weeks of the program for model training, only making predictions from the 26th week of a mother's pregnancy onward. Possible extensions of this work could explore hybrid methods, such as initially relying on predictions from other trained IBL models and gradually transitioning to the fully personalized model.

Acknowledgments

This research was supported by the AI Research Institutes Program funded by the National Science Foundation under AI Institute for Societal Decision Making (AI-SDM), Award No. 2229881

References

[1] S. O. Nogueira, A. McNeill, M. Fu, C. N. Kyriakos, U. Mons, E. Fernández, W. A. Zatoński, A. C. Trofor, T. Demjen, Y. Tountas, et al., Impact of anti-smoking advertising on health-risk knowledge and quit attempts across 6 european countries from the eurest-plus itc europe survey, Tobacco Induced Diseases 16 (2018). doi:10.18332/tid/96251.

[2] A. Lalan, S. Verma, P. R. Diaz, P. Danassis, A. Mahale, K. M. Sudan, A. Hegde, M. Tambe, A. Taneja, Improving health information access in the world's largest maternal mobile health program via bandit algorithms, in: Proceedings of the AAAI Conference on Artificial Intelligence, volume 38, 2024, pp. 22913–22919. doi:https://doi.org/10.1609/aaai.v38i21.30329.
[3] S. Amagai, S. Pila, A. J. Kaat, C. J. Nowinski, R. C. Gershon, Challenges in participant engagement and retention using mobile health apps: literature review, Journal of medical Internet research 24 (2022) e35120. doi:10.2196/35120.
[4] J. A. Killian, A. Biswas, L. Xu, S. Verma, V. Nair, A. Taneja, A. Hegde, N. Madhiwalla, P. R. Diaz, S. Johnson-Yu, et al., Robust planning over restless groups: engagement interventions for a large-scale maternal telehealth program, in: Proceedings of the AAAI Conference on Artificial Intelligence, volume 37, 2023, pp. 14295–14303. doi:https://doi.org/10.1609/aaai.v37i12.26672.
[5] P. Whittle, Restless bandits: Activity allocation in a changing world, Journal of applied probability 25 (1988) 287–298. doi:10.2307/3214163.
[6] P. Danassis, S. Verma, J. A. Killian, A. Taneja, M. Tambe, Limited resource allocation in a non-markovian world: the case of maternal and child healthcare, arXiv preprint arXiv:2305.12640 (2023). doi:https://doi.org/10.48550/arXiv.2305.12640.
[7] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural computation 9 (1997) 1735–1780.
[8] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, Advances in neural information processing systems 30 (2017).
[9] C. Gonzalez, J. F. Lerch, C. Lebiere, Instance-based learning in dynamic decision making, Cognitive Science 27 (2003) 591–635. doi:https://doi.org/10.1207/s15516709cog2704_2.
[10] E. H. Bugbee, C. Gonzalez, Making predictions without data: How an instance-based learning model predicts sequential decisions in the balloon analog risk task, in: Proceedings of the annual meeting of the cognitive science society, volume 44, 2022.
[11] S. B. Taieb, R. J. Hyndman, et al., Recursive and direct multi-step forecasting: the best of both worlds, volume 19, Department of Econometrics and Business Statistics, Monash Univ., 2012.
[12] T. N. Nguyen, C. Gonzalez, Theory of mind from observation in cognitive models and humans, Topics in cognitive science 14 (2022) 665–686. doi:https://doi.org/10.1111/tops.12553.
[13] J. R. Anderson, C. J. Lebiere, The atomic components of thought, Psychology Press, 2014. doi:https://doi.org/10.4324/9781315805696.
[14] C. Gonzalez, V. Dutt, Instance-based learning: integrating sampling and repeated decisions from experience., Psychological review 118 (2011) 523. doi:https://doi.org/10.1037/a0024558.
[15] ARMMAN, mmitra - armman - helping mothers and children, 2024. URL: https://armman.org/mmitra/.
[16] A. Mate, L. Madaan, A. Taneja, N. Madhiwalla, S. Verma, G. Singh, A. Hegde, P. Varakantham, M. Tambe, Field study in deploying restless multi-armed bandits: Assisting non-profits in improving maternal and child health, in: Proceedings of the AAAI Conference on Artificial Intelligence, volume 36, 2022, pp. 12017–12025. doi:https://doi.org/10.1609/aaai.v36i11.21460.
[17] Y. Zhao, T. Wang, D. Nagaraj, A. Taneja, M. Tambe, The bandit whisperer: Communication learning for restless bandits, arXiv preprint arXiv:2408.05686 (2024). doi:https://doi.org/10.48550/arXiv.2408.05686.
[18] G. Xiong, J. Li, Finite-time analysis of whittle index based q-learning for restless multi-armed bandits with neural network function approximation, in: Thirty-seventh Conference on Neural Information Processing Systems, 2023.
[19] N. Jaques, A. Lazaridou, E. Hughes, C. Gulcehre, P. Ortega, D. Strouse, J. Z. Leibo, N. De Freitas, Social influence as intrinsic motivation for multi-agent deep reinforcement learning, in: International conference on machine learning, PMLR, 2019, pp. 3040–3049.
[20] N. Modi, P. Mary, C. Moy, Transfer restless multi-armed bandit policy for energy-efficient heterogeneous cellular network, EURASIP Journal on Advances in Signal Processing 2019 (2019) 1–19. doi:https://doi.org/10.1186/s13634-019-0637-1.
[21] Y. Zhao, N. Behari, E. Hughes, E. Zhang, D. Nagaraj, K. Tuyls, A. Taneja, M. Tambe, Towards

zero shot learning in restless multi-armed bandits, AAMAS (2023). doi:https://doi.org/10.48550/arXiv.2310.14526.
[22] N. Behari, E. Zhang, Y. Zhao, A. Taneja, D. Nagaraj, M. Tambe, A decision-language model (dlm) for dynamic restless multi-armed bandit tasks in public health, arXiv preprint arXiv:2402.14807 (2024). doi:https://doi.org/10.48550/arXiv.2402.14807.
[23] S. Verma, Y. Zhao, S. Shah, N. Boehmer, A. Taneja, M. Tambe, Group fairness in predict-then-optimize settings for restless bandits, 2024.
[24] E. Lee, M. S. Lavieri, M. Volk, Optimal screening for hepatocellular carcinoma: A restless bandit model, Manufacturing & Service Operations Management 21 (2019) 198–212. doi:https://doi.org/10.1287/msom.2017.0697.
[25] T. Ayer, C. Zhang, A. Bonifonte, A. C. Spaulding, J. Chhatwal, Prioritizing hepatitis c treatment in us prisons, Operations Research 67 (2019) 853–873. doi:https://doi.org/10.1287/opre.2018.1812.

A. Real-world ARMMAN Dataset

A.1. Secondary Analysis

Our experiment falls into the category of secondary analysis of the data shared by ARMMAN. This paper does not involve deploying the proposed algorithm or any other baselines to the service call program. As noted earlier, the experiments are secondary analyses with approval from the ARMMAN ethics board.

A.2. Consent and Data Usage

Consent is obtained from every beneficiary enrolling in the NGO's mobile health program. The NGO owns the data collected through the program and only the NGO is allowed to share data. In our experiments, we use anonymized call listenership logs to calculate empirical transition probabilities. No personally identifiable information (PII) is available to us. The data exchange and usage were regulated by clearly defined exchange protocols, including anonymization, read-access only to researchers, restricted use of the data for research purposes only, and approval by ARMMAN's ethics review committee.

A.3. Interventions in ARMMAN data

The interventions in ARMMAN data are chosen based on restless multi-arm bandit (RMAB) algorithms [16, 17]. RMABs are a model for sequentially distributing scarce resources to a set of agents [5, 18, 19]. Concretely, we have a set of arms and a limited budget and face the question of deciding which arms to pull in each round. The state of arms evolves according to a Markov Decision Process where transition probabilities depend on whether the arm is pulled in this step. RMABs have a broad range of applications, including resource allocation in anti-poaching, machine maintenance, cellular networks [20, 21]. RMABs have had extensive use in healthcare settings such as call scheduling in a maternal and child care program [22, 23], screening patients at risk of cancer [24], and allocating hepatitis C treatment [25].

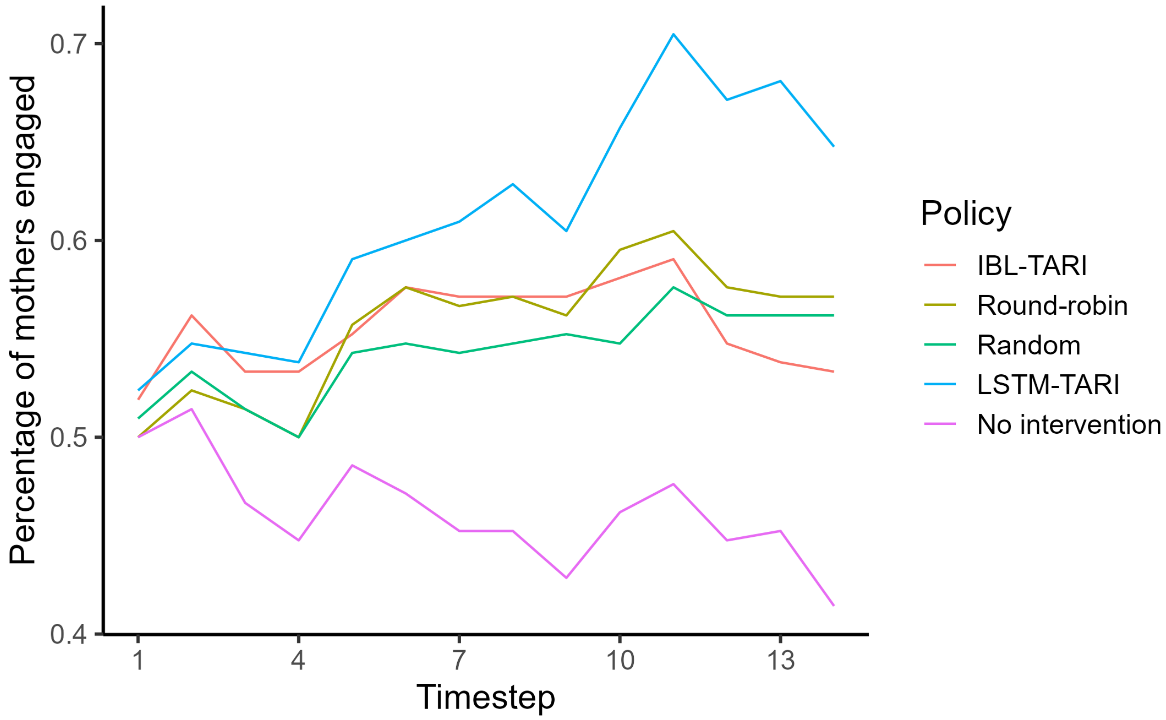

Figure 7: Percentage of mothers engaged in the program per testing timestep using simulated counterfactuals. $N = 210$, budget $k = 6$.