Computational Methods for Personalizing Mobile Health Interventions

by

Mawulolo Koku Ameko

A dissertation submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
at the
School of Engineering and Applied Science
University of Virginia
Charlottesville, VA

December 2021

Committee Approval:

We, the undersigned committee members, certify that we have reviewed this dissertation and approve it in partial fulfillment of the requirements of the degree of Doctor of Philosophy in Systems and Information Engineering.

Laura E. Barnes, Ph.D. (School of Engineering and Applied Science) Date
Advisor

Peter Beling, Ph.D. (School of Engineering and Applied Science) Date
Committee Chair

Mehdi Bouckechba, Ph.D. (School of Engineering and Applied Science) Date
Committee Member

Jundong Li (School of Engineering and Applied Science) Date
Committee Member

Bethany Teachman, Ph.D. (School of Art & Sciences) Date
Committee Member

Certified by:

Jennifer L. West, Dean, School of Engineering and Applied Science Date

Abstract

The increasing use of smart devices such as smartphones and wearables has enabled new opportunities for health research that leverage rich multimodal, multiscale data streamed from embedded sensors to monitor and deliver timely interventions to patients when and where they need them most using recent computational advances in machine learning. This new approach to intervention provides more accessible, scalable, and cost-effective options to reach individuals. This relatively new interven -tion framework called just-in-time adaptive intervention (JITAI) aims to provide the right type/amount of support, at the right time, by adapting to an individual's dynamically changing internal and contextual state. The success of JITAIs depends on accurate models for recognition of internal states such as an individual's emotional state and other contextual states relevant to health. This data can in turn be used to design an intervention policy that leads to improved user engagement, lower attrition rates, and lower symptom burden. In this dissertation, we propose multiple computational techniques that move us towards more personalized JITAIs for mental health.

To demonstrate the efficacy of our proposed computational approaches, we leverage real-world data from multiple mobile sensing studies from a population of college students to (1) personalize affect recognition for subgroups of individuals, (2) learn context-aware intervention policies for emotion regulation (ER), and finally culminating by combining approaches 1 and 2 into a (3) subgroup-based, context-aware intervention policies for emotion regulation. These methodologies contribute to a growing body of approaches that moves us closer to the realization of just-in-time interventions in mobile health.

Acknowledgments

I would like to thank a few people that have helped me along this long journey of completing my doctoral studies.

First of all, I would like to thank my advisor Laura Barnes for giving me the opportunity to study under her supervision and for never falling short of supporting me with my development as an academic. I especially want to thank her for being very patient with me all along the way. I would like to thank Mehdi Boukhechba for the energy and push he provided towards my publications over the past 4 years. And thank you to Bethany Teachman for providing a lot of timely feedback for the success of this dissertation. I also say a big thank you to Jundong Li and Peter Beling for being available to provide feedback on my research ideas.

I am grateful to my wonderful lab members for making my time in the lab worthwhile. I want to especially thank Lihua Cai for being such a wonderful collabo -rator; I cannot forget the several nights of calls to discuss research ideas and write papers for conference deadlines.

I want to thank Charlottesville Christian Fellowship for providing me with the support and grounding I needed through the highs and the lows. You have been closer than family from the very first day! I want to thank my sliblings, Eyram, Mawuli and Jonathan for always providing a smile when I needed and showing me that no matter what, we have each other. And to my parents, Kouma and Charity, I love you and I hope I have made you proud. And yes, dad, I will add the "rev" at some point!

To my wife, Ilone, words will fail me to express my gratitude to you. Thank you for so graciously putting up with my work life situation over the past two years. You give me a reason to work hard, and I hope I have made you proud. Cheers to more years of love and growth while we stand firm upon the solid rock of Christ.

Finally, I thank God for giving the ability and grace to come this far.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

CMAB	Contextual Multi-Armed Bandits
EMA	Ecological Momentary Assessment
EMI	Ecological Momentary Interventions
ER	Emotion Regulation
GPS	Global Positioning System
JITAI	Just-In-Time Adaptive Intervention
MAB	Multi-Armed Bandits
MDP	Markov Decision Processes
OSM	OpenStreetMap
RL	Reinforcement Learning
SAMMI	Social Anxiety Monitoring and Mobile Intervention Study
SIAS	Social Interaction Anxiety Scale
SMS	Short Messaging Service
SVM	Support Vector Machines

1 | Introduction & Related Work

1.1 Motivation

The ubiquity of smart devices (e.g., mobile phones, smartphones, laptops, and tablets) in recent years has created massive opportunities for monitoring human health traces from sensors (e.g., GPS, accelerometer, microphone etc) and delivering timely interven -tions to reinforce health habits with potential positive impacts on long-term mental health and wellbeing. There is a wide range of application areas using mobile health interven -tions including, physical activity, alcohol use, smoking cessation and mental illnesses. This new technology allows us to innovate the traditional approach of delivering treatment in clinic, which typically involves regular visits to the clinic for treatments. This approach to mental health treatment is inefficient and not easily accessible. Furthermore, it is also fraught with issues of recall bias on the side of the patient not being able to accurately remember events that happen between visits. In recent years, new approaches to mitigate the effect of recall bias have been developed to capture natural experiences of patients in their natural environments using surveys delivered periodically through smart devices that enables patients to log their experiences. This methodology is referred to as experience sampling methodology or ecological momentary assessment (EMAs) [Shiffman et al., 2008].

EMAs hold the promise of collecting high quality data for behavioral mental health research to enable statistical inferences from both an idiographic and nomothetic perspective. When EMAs involve active interventions when a risk of interest is detected, it is called an ecological momentary interventions (EMI). The capability of

EMA or EMI studies as an efficient alternative to supplant the traditional approach is often undermined by issues such as user attrition, confounding bias, habituation to EMA prompts and perceived relevance of interventions [Doherty et al., 2020]. These issues have spurred new research directions to develop methods geared towards reducing these issues that plague the efficacy of EMA studies. One line of research uses gamification to improve user engagement for EMA studies [Doherty et al., 2020, Klasnja et al., 2015], but not much has been done to improve the relevance of interventions (EMIs). It is important that systems built to deliver interventions to patients be personalized and effective especially in the early stages, to mitigate the risk of user engagement for lost of trust in the system, or the exposure to unexpected adverse outcomes caused by poor interventions [Tewari and Murphy, 2017]. This is especially relevant in the mental health application settings where the negative impacts can be devastating [Doherty et al., 2020]. In the traditional recommender systems literature, the problem of recommending relevant actions to users with limited data at early stages is called the cold-start problem.

In this thesis, we develop computational methods similar to cold-start approa -ches to improve the relevance of mobile health interventions by means of personalized health assessments and intervention recommendations. Our methods were developed with purely observational data to improve the relevance of mobile interventions especia -lly at the early stages of an EMA study, which consequently will impact user engage -ment and produce desirable outcomes for users. The framework to operationalize our methods is called the just-in-time adaptive interventions (JITAI).

1.2 Methods for Personalization

The concept of personalization is widely used in more traditional recommender systems such us e-commerce and entertainment systems like Netflix to improve user engagement and relevance scores. Models such as collaborative filtering, matrix factorization and deep learning form the cornerstone of the immense methodological advances in recommender systems. More recently, there has been new inroads made with non-

traditional methods such as contextual bandits and causal inference to improve upon the personalization efforts to make recommendations more adaptive over time via safe exploration and exploitation strategies [Goldenberg et al., 2021]. There is a growing effort in mobile health to adopt these techniques for more nuanced applications like mental and physical health which have relatively more consequential implications beyond user disengagement; for example, providing unhealthy recommendations that could lead to a patient relapse in an alcohol cessation study. Papers focusing on personalization in mobile health can be mainly categorized by two modeling targets; symptoms predictions and intervention recommendations.

There have been several papers showing improvements in training personalized, that is individual-level, models over a more generalized model in mobile health symp -toms. For example, Jacques et. al, showed that by using multi-task learning, where the task is defined at the individual level, leads to improvements in predicting health, stress and well-being over a more generalized model [Jaques et al., 2016]. Also, Koldijk et al. [Koldijk et al., 2016] showed that by adding the participant ID as a feature to their workplace stress prediction model improved accuracy in classifying mental effort suggesting that person-level models are more useful. An even more recent paper proposed a collaborative filtering based classifier to detect depressive symptoms [Xu et al., 2021] and showed generalization between data sets collected from different institutions, thus highlighting the importance of personalized models for study replication. Similarly, for mobile health interventions optimization, there is a growing interest in using contextual bandits to personalize interventions based on users' contexts. A notable work by Rabbi et. al. used contextual bandits to personalize physical activity recommendations to users in a study conducted over 14 weeks [Rabbi et al., 2015]. Tomkins et. al. also demonstrated the use of contextual bandit with multi-task learning to intelligently pool data from similar users for the recommendation of physical steps; their findings showed significant improvements in performance over a more generalized contextual bandit approach [Tomkins et al., 2021]. Throughout this thesis, we develop techniques for personalization in mobile health and demonstrate significant advances over more generalized models.

1.3 Just-in-Time Adaptive Interventions

1.3.1 Mobile Health Interventions

Many existing papers propose recommender systems targeting different health outcomes. For example, myBehavior, a mobile app that tracks user's physical and dietary habits, recommends personalized suggestions for a healthier lifestyle [Rabbi et al., 2015]. Cheung et al. [Cheung et al., 2018] created a mobile app called IntelliCare, which consists of a suite of 12 individual apps as 'treatments' that will be recommended for managing depression and anxiety. Yang et al. [Yang et al., 2018] created a mobile health recommender system that integrates depression prediction and personalized therapy solutions to patients with emotional distress. In their system, personalization is realized using 9 external factors related to depression, including family life, external competition, interpersonal relationship, self-promotion burden, economic burden, work pressure, individual personality, coping style, and social support, which are assessed using mobile questionnaires. These mobile health efforts are consistent with a mobile intervention framework called Just-in-time adapti -ve intervention (JITAI) [Nahum-Shani et al., 2017].

1.3.2 Reinforcement Learning Powered Mobile Health

Two aspects regarding the intervention decisions made in a JITAI framework are the timing of intervention delivery and choosing the best intervention strategy to deliver. Most existing work focuses on optimizing for the best timing to deliver an intervention (e.g., predicting stressful moments linked to emotional eating [Rahman et al., 2016]). By contrast, our work focuses on identifying the most effective ER strategies based on a person's context (e.g., location, activity). Reinforcement learning with Markov decision processes (MDPs) are typically used to operationalize the key objectives of a JITAI. Example applications include personalizing sepsis treatment strategies [Peng et al., 2018], encouraging physical activity for diabetes patients [Yom-Tov et al., 2017], and managing stress [Jaimes et al., 2014]. Interestingly, although reinforcement

learning is not directly applied to recommend ER strategies for emotion management, it has been applied to understand the psychological and cognitive process of ER [Marinier et al., 2008, Raio et al., 2016]. Specifically, [Marinier et al., 2008] argues that emotions are a proxy for human subject decision making. In other words, humans make decision to maximize the way the feel afterwards. Similarly, [Raio et al., 2016] makes the argument that emotion regulations do not only fit the computational model of reinforcement learning using a simple classification of model-free or model-based learning, but argues for a more hybrid and hierarchical approach.

1.3.3 Learning Initial Policies for Fast Online Adaptation

As the adaptation of reinforcement learning become more widespread in applications including autonomous vehicles, healthcare and robotics; there has been a growing body of research exploring safe exploration in RL or other offline alternatives in settings like clinical healthcare where active exploration is unethical or dangerous. Some notable examples include learning vasopressor and IV fluid dosages for sepsis management [Raghu et al., 2017] using Double-Deep Q-Network [Van Hasselt et al., 2015] with doubly robust policy evaluation [Jiang and Li, 2016], and the AI clinican [Komorowski et al., 2018] also using deep reinforcement learning for sepsis treatment in the intensive care unit setting. These examples however represent settings where active exploration from reinforcement learning is prohibitive or impossible. On the other hand, in more moderate stake applications such as robotics, research has been developing in recent years to enable offline policy learning that uses abundant historical data of robots to learn better policies to prevent the wear and tear on the robot and enable fast adaptation. In fact, several recent benchmarks have been developed to advance research in this area, notable RL Unplugged [Gulcehre et al., 2020] and D4RL [Fu et al., 2020]. Recent work, showed the benefit of learning an offline policy for fast adaptation in the online setting [Rakelly et al., 2019] in robotic applications. While researchers in mobile health indicate that there might be similar benefits for offline learning in mobile health settings through learning an initial policy [Tewari and Murphy, 2017], there has been no real world demonstration of the approach to

our knowledge. It is, however, worth noting there is recent theoretical work in offline reinforcement learning where the average undiscounted reward is optimized [Liao et al., 2020b, Liao et al., 2020c]. We will focus on the practical demonstration of offline learning with purely observational data in this thesis work motivating future work to evaluate our approaches in a real-world setting.

1.4 Preliminary Work in Reinforcement Learning

In general, we denote random variables as capital letters and their realizations in small letters (e.g., x is a realization of X random variable). Let us denote S, A, R as the random variables for state, action, and associated reward for the sequential decision making process. To set up the mathematical formulation for mobile health interventions, it is assumed that we collect multimodal data streams including states (s), actions (a) and rewards (r) from a micro-randomized trial [Klasnja et al., 2015] over a period T. Consequently, we obtain data, denoted $D = \{(s_t, a_t, r_t)\}_{t=1}^{T}$. Typically, the sequence of data is assumed to be generated from a Markov decision process, with the implication that expected reward at each point is sufficiently determined by the current state given the immediate past state. In mobile health intervention applications, it is assumed that T can take infinite values which makes the problem an infinite horizon Markov decision process.

Markov Decision Process (MDP)

Suppose we observe a training dataset, $D_n = \{D_i\}_{i=1}^{n}$ that consists of n independent, identically distributed (i.i.d.) observations of D:

$$\{S_1, A_1, S_2, \ldots, S_T, A_T, S_{T+1}\}$$

where t indexes the decision time. The length of the trajectory is assumed to be non-random (e.g., 5 weeks for Social Anxiety Monitoring and Mobile Intervention (SAMMI) data). $S_t \in S$ is the state at time t and $A_t \in A$ is the action (treatment/strat

-egy) selected at time t. We assume the action space, A, is finite. Without loss of generality, we assume that state space, S, is finite; as this imposes no practical limitations and can be extended to a general state space.

The states evolve according to a time-homogeneous Markov process. For $t \geq 1$, $S_{t+1}\{S_1, A_1, \ldots, S_{t-1}, A_{t-1}|S_t, A_t\}$, and the conditional distribution does not depend on t. Denote the conditional distribution by P, i.e., $Pr(S_{t+1} = s^{'}|S_t = s, A_t = a) = P(s^{'}|s, a)$. The reward (i.e., outcome) is denoted by $R_{t+1} = \mathcal{R}(S_t, A_t, S_{t+1})$. We assume the reward is bounded, i.e., $|\mathcal{R}(s, a, s^{'})| \leq R_{max}$. We use $r(s, a) = \mathbb{E}[R_{t+1}|S_t = s, A_t = a]$.

Let $H_t = \{S_1, A_1, \ldots, S_t\}$ be the history up to time (t - 1) together with the current state, S_t. Denote the conditional distribution of A_t given H_t by $\pi_{b,t}(a|H_t) = Pr(A_t = a|H_t)$. Let $\pi_b = \{\pi_{b,1}, \ldots, \pi_{b,T}\}$. This is called the behavior policy in the literature. In this work, we are not required to know the behavior policy.

Consider a time-stationary, Markovian policy, π, that takes the states as input and outputs a probability distribution on the action space, A, that is, $\pi(a|s)$ denoting the probability of selecting action, a, at state, s. The average reward of the policy,π, is defined as

$$\mathbb{V}(s|\pi) = \lim_{T \to +\infty} \mathbb{E}_\pi \left(\frac{1}{T} \sum_{t=1}^{T} \gamma R_{t+1} | S_1 = s \right), \tag{1.1}$$

where the expectation, $\mathbb{E}_\pi$, is with respect to the distribution of the trajectory in which the states evolve according to P and the actions are chosen by π. Also, the parameter $\gamma \in [0, 1]$ is added to encode the idea that immediate rewards are worth more than long-term reward, although, recent literature shows the irrelevance of the γ in mobile health applications [**?**]. It can be shown that the maximal average reward among all possible history dependent policies can be in fact achieved by some time-stationary, Markovian policy. Consider a pre-specified class of such policies, Π, that is parameterized by $\theta \in \Theta \subset R^p$. We state the objective of learning as an optimization problem;

$$\pi^* = \underset{\pi \in \Pi}{\operatorname{argmax}}\, \mathbb{V}^{\Pi}(s|\pi). \tag{1.2}$$

1.4.1 Limitations as Applied to Mobile Health

Using Markov decision processes with reinforcement learning (RL) seems keenly appro -priate for mobile health data. However, there are a few weaknesses that limit the utility in real-world applications.

Impact of Noisy Observations

Mobile health intervention studies are mostly conducted in uncontrolled environments; both context information and rewards can be very noisy as a result of possible confounders. For example, in the SAMMI data, the effect of strategies is represented by self-reported effectiveness of affect, but this measurement is subjective and could be limited by recall bias. Similarly, sensor pedometer step count data could be confounded by accidental hand movements. In addition, these data do not completely describe the context of the user. Consequently, such uncertainty typically requires even more interaction data to select an optimal policy.

Although it is typically useful to optimize for long-term effect of mobile health policies to account for the risk of habituation; this approach often leads to high variance in reward estimation as well as a slow learning. Specifically, by setting the discount rate γ in 1.1 close to 1, an infinite amount of data is required for learning [Arumugam et al., 2018, François-Lavet et al., 2015, Jiang et al., 2015]. In effect, reducing the discount factor to be closed to 0 mitigates the risk of overfitting [Arumugam et al., 2018], the richness of the policy class to learn the optimal policy depends not on the state and actions but on the planning horizon denoted above as T, that is the shorter the horizon is, the more likely it is to learn a good policy and vice versa. In light of these limitations, most works applying reinforcement learning in real world mobile health intervention applications, tend to use bandits or contextual bandits since these can be seen as a full RL with a discount factor $\gamma = 0$.

Reliance on Access to True Underlying State

One other weakening assumption for using full RL with MDPs is the underlying assumption of having access to the true state space of an agent or a user in a mobile health setting. This is almost never realized in practice since as stated above, the contextual variables are a noisy representation of the true underlying state of a user. Consequently, the impact of the Markov assumption is aggravated since the current state is not even well identified with available data. A recent paper, which deployed a real world application of mobile health interventions to encourage physical activity [Yom-Tov et al., 2017], argued that while there are methods using Q-learning or TD-learning for RL, these methods heavily rely on the assumption of having access to the true underlying state, or to high-quality features that represent the dynamics well. As a result, they used a contextual bandits which allows them to predict the immediate effect of a given interventions as opposed to additionally changing the state of a user. There are several other studies, that made similar arguments to select contextual bandits or some variant thereof as opposed to a full-blown RL, see [Rabbi et al., 2015, Paredes et al., 2014, Forman et al., 2019]. Similar to previous work, we propose to use contextual bandits to allow us to learn about the immediate effect of treatments/strategies from the limited historical data from the SAMMI study. Our objective, is to use the ensuing policy as a warm-start treatment assignment policy to collect high quality data less prone to the risk of disengagement and attritions in a future study.

1.4.2 Contextual Bandits

Contextual multi-armed bandit (CMAB) or simply contextual bandits is an reinforce -ment learning framework that leverages contextual information to learn a policy that triggers actions based on the context to achieve optimal expected rewards. Typically, CMAB consists of an agent that interacts with an environment over a finite number of trials $i = 1, 2, \ldots, T$ such that: 1) it observes a context x from an input space $\mathcal{X}$; 2) chooses an action from a set $\mathcal{A} = \{a_1, a_2, \ldots, a_{k-1}, a_k\}$, which contains all the

strategies that each corresponds to an arm of a k MAB; and 3) receives a reward signal r_i. The goal of the agent is to learn a policy to guide action decisions. Unlike a full-blown reinforcement learning algorithm typically modeled using MDPs, where an action decision modifies future states and action selections, CMAB assumes that $\{(x_t, a_t, r_t)\}_{t=1}^{T}$ are independently and identically distributed following an unknown generative distribution $\mathcal{D}$.

The observables are $(x_t, a_t, r_t(a_t))$; in particular, only the reward $r_t(a_t)$ for the chosen arm a_t is observed. For each context $x \in \mathcal{R}^d$, the optimal assignment is $a(x) = \text{argmax}_a \mathbb{E}[r(a)|x]$ and we let $a_t = a^{(x_t)}$, which denotes the optimal assignment for context x_t. The objective is to find an assignment rule that sequentially assigns a_t to minimize the cumulative expected regret $\sum_{t=1}^{T} E[r(a_t) - r(a^t)]$, where the assignment rule is a function of the previous observations $(x_j, a_j, r(a_j))$ for $j = 1, ..., t-1$ and of the new context x_t.

Contextual bandits typically assumes an active learning setting where an agent iteratively learns to make good decisions by interacting with the environment through experimentation. This is however costly and expensive for most real world safety-critical applications such as healthcare or mobile health where there is a risk of taking dangerous decisions or causing a disengagement from the users. Consequently, this motivates a new paradigm of reinforcement learning in the offline setting, also known as, offline reinforcement learning or batch reinforcement learning. In this setting, the agent does not explore the environment, instead, it has access to log of historical trajectories from other agents' behaviors and the goal is to learn an optimal policy from this data. As stated above, these historical data pose a missing data problem, in that, for each context the agent only observes the reward associated with the action taken and not the reward of the alternative actions. In other words, there is an abundance of factual outcomes of actions, but the agent needs to estimate the counterfactual outcomes with limited data to estimate the associated reward of the alternative actions. This are concept is well established from the causal inference body of literature.

Causal Inference

Causal inference fundamentally aims to answer the cause-and-effect question: does X cause Y? If so, how much is the effect of X on Y? Causal inference helps us learn about how things work and predict the impact of a change on an event of interest [Morgan and Winship, 2015, Imbens and Rubin, 2015]. In this section, we review some basic concepts and assumptions of causal inference that will be used in the later chapters of this dissertation when we make a connection between causal inference and health treatment recommendation.

The Potential Outcome Framework

The potential outcome framework of causal inference [Rubin, 1974] is the most widely used formulation of causality. We use random variable $A = \{a_1, a_2, \cdots, a_k\}$ as a collection of actions an agent can take in an environment and assume that only one treatment can be taken each time, i.e., a_k is either 1 (when action is taken) or 0 (when action is not taken). In this framework, each individual has k potential outcomes $Y(a_k)$ depending on the value of a_k. For example, in the bandit setting, each time, we assume there is a potential outcome $Y(a_k = 1)$ if the action a_k is taken and $Y(a_k = 0)$ if not taken.

One measurement or estimand of the causal effect is the average difference (over individuals) between those potential outcomes. It is called the average treatment effect (ATE): $E[\delta] = E[Y(1)] - E[Y(0)]$, where the expectation is taken over the whole population of interest. In the language of graphical models [Pearl, 2009], this is framed as evaluating the impact of an intervention on random variables in a probabilistic graph. The difficulty of causal inference is that we can only observe one realization of all the potential outcomes $Y(a)$, for $a \in \{0, 1\}$. The counterfactual estimand of the value of a treatment when there is additional contextual data is called the conditional average treatment effect (CATE). For example, given a context x the CATE of treatment is denoted $CATE(a = 1|x) = Y(a = 1|x) - Y(a = 0|x)$.

Randomized Experiments and Observational Studies

There are two types of data commonly encountered in causal analysis: data collected from randomized experiments and data collected from observational studies. Randomi -zed experiments are studies where each unit receives a treatment randomly, for example, as in a clinical trial where a random proportion of patients receives treatment and placebo. This approach ensure reliability and validity of statistical estimates for causal effects. A naive ATE estimator of the difference between the treated and untreated with data from randomized experiments is unbiased. While such data is of great quality, it is rarely easy to obtain.

Observational data, on the other hand, is collected from studies where we have no control over the treatment assignment mechanism. This often results from studies where it is impractical to perform a randomized experiment (e.g., for ethical reasons), or when we cannot control the data collection process. Special care is required when making causal statement with observational data, since the naive ATE estimator is generally biased. Despite such difficulty, observational data is easily accessible compared to data collected from randomized experiments.

To estimate the CATE from observational data, the following set of assumptions are made;

- **Ignorability**. $\{Y_i(a_1), Y_i(a_2), \ldots, Y_i(a_k)\}A_i|X_i = x$ for any $x \in X$ and $a_i \in A = \{a_1, a_2, \cdots, a_k\}$. In other words, there are no unobserverd confounders.
- **Positivity or overlap**. That is that $0 < P(A = a_i|x) < 1$ for all $a_i \in \mathbb{A}$ and $x \in X$. In other words, there is a non-zero probability for selecting any action in a given context.

In summary, offline contextual bandits is equivalent to deriving the actions that maximize the CATE as follows: $a^*(x) = \text{argmax}_a\, CATE(a|x)$.

If we treat recommending a mobile health intervention to a user as action taken by an agent, the data collected from a lot of behavioral health intervention systems is an example of observational data. We leverage this connection in Chapters 3 and 4.

Ideally, conducting a live experiment, akin to a randomized control trial is required to learn the treatment effect of different strategies [Klasnja et al., 2015]. However, this approach is typically fraught with issues including cost of experimenta -tion. Experiments with human subjects to discover treatments are expensive and go through a lengthy process of approval, and even when finally staged take a significant amount of time to complete. However, there is an abundance of observational data collected from EMAs and smartphone embedded sensors, which can be leveraged to warm-start the process of treatment selection in mobile health intervention designs. These data are cheap to obtain; for example, in this thesis, we mostly use data from a study targeted toward identifying the treatment outcome of cognitive-bias modifications to learn about a somewhat related topic, namely, emotion regulation. Moreover, we can use observational data to learn about the contextual effectiveness of different treatments which are almost impossible in live human trials. In light of these merits, we propose a novel approach to the treatment of social anxiety that leverages smartphones and contextual bandits to build a initial or warm-start recommender algorithm for emotion regulation strategies tailored to a person's contexts as well as the idiographic characteristics.

1.5 Contributions

The contributions proposed in this thesis work are three-fold:

1. We develop, to our knowledge, the first methodological approach to learning an initial policy in the mobile health interventions setting using available retrospective mobile sensing data.

2. We develop a framework to learn a subgroup-based, personalized mobile health intervention policy in the offline setting.

3. We establish a proof-of-principle with a case study on emotion-regulation, an important and trans-diagnostic process that will likely have broad impacts on mental health.

1.6 Hypotheses

To achieve the objectives of this thesis, we investigate three hypotheses:

1. There is a subgroup-based prediction model that outperforms a generalized model for affect prediction tasks.
2. There is a general context-aware intervention policy for mental health which performs significantly better than what people are already doing.
3. There are subgroups in the given population for which there is an enhanced effectiveness of intervention effects based on contextual variables.

1.7 Organization

The rest of the thesis is organized as follows:

1. In Chapter 2, we leverage cluster-based approaches to predict human affective states from passive sensor data.
2. In Chapter 3, we use contextual bandits and causal inference techniques to demonstrate the learning and evaluation of mobile health interventions policies on the process of emotion regulation. We also, provide results from our extension of the approach to a more clinically relevant application.
3. In Chapter 4, we use subgroup analysis to learn warm-start policies of the emotion regulation intervention policy and demonstrate improvement on the approach outlined in Chapter 3.
4. In Chapter 5, we conclude the thesis by summarizing impact of the body of work in this thesis, the limitations as well as an outline of possible future directions.

2 | Personalized Affect Prediction

Negative affect is a proxy for mental health in adults. By being able to predict participants' negative affect states unobtrusively, researchers and clinicians will be better positioned to deliver targeted, just-in-time mental health interventions via mobile applications. In this chapter, we present our to personalizing the passive recognition of negative affect states via a group-based modeling of user behavior patterns captured from mobility, communication, and activity patterns. Our empirical experiments show that group models outperform generalized models in a dataset based on a two weeks of users' daily lives.

2.1 Introduction

The extent to which individuals experience positive and negative affect on a daily basis is associated with mental health outcomes [Clark et al., 1994]. Higher levels of negative affect are associated with increased vulnerability to many mental disorders, including depression and anxiety disorders, two of the most common types of mental disorders in U.S. adults [Kessler et al., 2005a]. Mental health research typically relies on self-report questionnaires that assess negative affect at a moment in time. Repeated administration of these measures, such as in an ecological momentary assessment (EMA) framework, is resource intensive and susceptible to retrospective bias when participants are asked to recall their mood over a previous duration [Gentzler and Kerns, 2006]. Ideally, negative affect would be recognized without asking partici -pants, thereby reducing burden, improving compliance among participants, and

allowing for continuous modeling of affect change. To aid recognition of negative affect, unobtrusive mobile sensing of location, texts and calls, and activity levels could also be used to enrich the information provided by participants' responses to questionnaires assessing negative affect and measures of mental health (e.g., social anxiety, depression).

Current affect recognition approaches are based primarily on generalized or individualized approaches [Yonekura et al., 2016]. In generalized approaches, the recognition model learns global patterns that the majority of participants followed during the experiment. These patterns are then used for prediction. Since user behaviors vary substantially, generalized models may fail to predict variations in affect for an individual person. In contrast, individualized models are designed to learn participants' patterns on a case-by-case basis, thus they are expected to be more accurate. However, individualized models require a certain number of observations for each individual to obtain robust prediction performance. In short-term studies involving human subjects (e.g., two weeks), individual models may fail to adequately capture individual affective patterns because of a small pool of observations [LiKamWa et al., 2013].

In our work, we propose a new group-based approach that integrates generalized and personalized models. We first propose a method for clustering multi-modal behavioral profiles that groups participants based on their mental states, activity levels, communications, and mobility patterns. We then apply several prediction algorithms to investigate whether group models using multi-modal user profiles outper -form the generalized or population-based model.

2.2 Related Work

Smartphone usage can be used as an indirect marker of mood. Passively sensed location information has been used to predict depressive symptoms [Saeb et al., 2015]. Individuals with higher social anxiety levels were more likely to report negative affect during the day, which in turn was predictive of spending more time at home at

subsequent measurements [Chow et al., 2017]. Self-reported stress and mental health indices were also successfully predicted in a 10-week long study design in college students with both passively and actively sensed data [Wang et al., 2017].

Prediction of affect from mobile sensing appears to be more difficult to replicate. In a feasibility study, LiKamWa et al. [LiKamWa et al., 2013] explored a personalized feature selection approach to predict changes in mood from unobtrusively sensed indices of social activity (e.g., calls/texts, emails), physical activity (e.g., GPS), and general mobile phone use (e.g., application use, web browsing). The study relied on two months of data collected from 32 participants. Results indicated high levels of accuracy in predicting mood using personalized models. The personalized modeling also produced better accuracy compared to a generalized model using data from all users.

A follow-up study in which a personalized feature selection approach was used to predict affect ratings from 27 participants over 42 days found no clear benefits of using this approach [Asselbergs et al., 2016]. However, these studies did differ in length, participant variability (e.g., depressive symptoms), and unobtrusive features assessed. It remains possible that personalized feature selection requires an intensive level of data collection that participants may perceive as burdensome. Given these findings, we use an intermediate approach between generalized and personalized models to recognize affect in a given situation.

2.3 Study Design

Sixty-five undergraduate students were recruited for a two-week study period to understand dynamics of emotional, cognitive, and interpersonal processes associated with depression and social anxiety. University students provide a relatively homoge -neous sample in terms of life phase and common psychological stressors, thereby mitigating the impact of a wide variety of potential nuance factors. Pre-study surveys were given to the students at enrollment, and one of these surveys measured students' social anxiety (SIAS) [Mattick and Clarke, 1998]. The study contained an ecological

momentary assessment (EMA) phase that requests self-report data on psychological affect throughout the day. A customized mobile app (Sensus) [Xiong et al., 2016] was installed on participants' personal Android smartphones and was programmed to deliver 6 EMAs throughout the day (each survey contained 12 questions), randomly scheduled in each 2-hour block from 9 a.m. to 9 p.m. (e.g., once between 9-11 a.m., once between 11 a.m.-1 p.m., etc.). Sensus was also configured to deliver an end-of-day survey at 10 p.m. each day. Prompts concerning affect first asked participants to rate how positive they were feeling from 1 (not at all) to 100 (very positive). The second question asked participants to rate how negative they were feeling from 1 (not at all) to 100 (very positive). In addition to these active assessments, Sensus also passively collected GPS coordinates every 150 seconds and accelerometer data at 1 Hz, in addition to call and text logs. All data were transmitted wirelessly to a secure Amazon Web Services server, where data were stored for further analysis (see Figure 2.1).

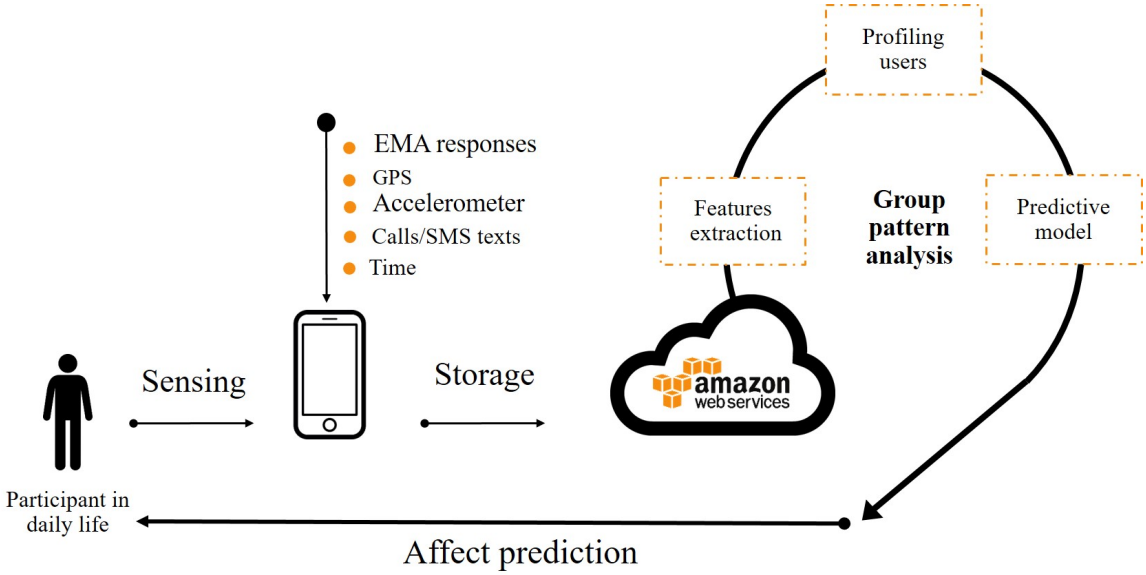


Figure 2.1: Passive and affect data collection using smartphones.

2.4 Experiments

2.4.1 Data Preprocessing

We first processed participants' raw GPS data into semantic locations (e.g., leisure, education, and home) by combining a spatiotemporal clustering algorithm [Kang et al., 2005] and OpenStreetMap (OSM) geodatabase [Keßler, 2015]. Our label taxonomy includes the following types: Education (e.g., university and libraries), Leisure (e.g., restaurants and cinemas), Out of town, In transition (e.g., going from one place to another), Home, and Other houses. Our algorithm has been trained to recognize Home as the place having a house OSM-tag (e.g., apartment, dormitory, house, etc. See [Keßler, 2015] for more details about OSM tags) where a subject stayed the most between 10 p.m. and 9 a.m.

For accelerometer data, we used statistical measures (mean, minimum, maximum, standard deviation, median and variance) on the 1-minute sliding window to extract several features of phones' motion around affect assessment moments. These features aim to represent the physical activity levels of the participants, and we used them to predict momentary negative affect. Note that our accelerometer features are extracted from the magnitude of acceleration $\sqrt{\frac{x^2+y^2+z^2}{3}}$ to make them orientation free, since the phones were used in participants' natural environments.

Individuals' affect may be associated with the degree to which they interact with others. Thus, we included communication events in our models. For each EMA we collected the number of text messages and phone calls as long as their duration overlapped with epochs prior to the EMA prompt. Here we chose 1 hour prior to the EMA prompt as the time window to record the number of text messages and phone calls.

2.4.2 Profiling Users

After preprocessing the data, we clustered the participants based on their behavioral profiles. There are different ways to cluster participants. For instance, a clustering

strategy can be based on time spent at home to cluster people having depressive symptoms, drawing on the hypothesized correlation between home staying and affect fluctuation patterns. The following four passively sensed profile features were used to drive the clustering process.

Location

For location data, we considered five common point-of-interest classes consisting of {'out of town', 'education', 'friends' houses', 'home', 'leisure'}. Then we calculated the proportion of time spent in each of these locations over the study period for each participant.

Activity

From the accelerometer data, we chose thresholds of 0.2 and 0.3 between the minimum and maximum to define three levels of activity (e.g., {Low, Medium, High} in acceleration). We chose these cutoffs based on the observed distribution of the acceleration values. Then for each participant, we calculated the proportion of time being in these activity levels (e.g., proportion of time being in the high level).

Short-Message Service (SMS)

From the SMS data, we aggregated the number of text messages sent and received within each 1-hour window during the study period. From this, we defined 5 text messaging levels based on text message frequencies (e.g., 'VeryLow','Low','Medium','High','VeryHigh') with intermediary cutoffs at 1, 10, 20 and 30 messages per hour based on their observed distribution.

Phone Calls

Similarly, we computed the proportion of calls occurring at each level of call activity defined as 'Low','Medium','High','VeryHigh' using thresholds of 1, 3 and 6 calls per 2-

hour window. We used a 2-hour window to accommodate the lower hourly frequency of phone calls compared with text messages.

Formally, for the design matrix $X \in \mathbf{R}^{N \times d}$ with $X = \{\mathbf{x_i}\}_i^N$, the feature vector for each participant is $\mathbf{x_i} = \{\underbrace{x_{i1}, x_{i2}, \cdots, x_{ip_1}}_{\text{M1}}, \underbrace{x_{i1}, x_{i2}, \cdots, x_{ip_2}}_{\text{M2}}, \cdots, \underbrace{x_{i1}, x_{i2}, \cdots, x_{ip_n}}_{\text{Mn}}\}$. Note that M_i $(i \in [1, n])$ represents the ith modality and p_i the number of levels in the ith modality.

With the above, we determine different clusters based on various combinations of these four passively sensed modalities in addition to SIAS using the G-means (Gaussian Means) [Hamerly and Elkan, 2004] algorithm. The G-means algorithm is an extension of K-means where number of clusters is automatically determined by iteratively selecting k such that the data assigned to each cluster follows a Gaussian distribution.

2.4.3 Predictive Models

We used 4 algorithms to test the predictability of negative affect: Gaussian process, SVM, linear lasso, and random forest. Each of these models has merit with respect to the issues that may ensue from constraints of data availability for model training, which is the case in this study. Although random forest, SVM, and Lasso regression are well-studied, Gaussian processes have demonstrated promising performance in e-health applications [Clifton et al., 2013] mostly because they enable experts to encode their beliefs about smoothness or periodicity using covariance functions. In addition, the complexity of the model is inherently regulated (see chapter 5 of [Rasmussen and Williams,]) and provides uncertainty over predicted values. In our case, we used the squared-exponential covariance function [Rasmussen and Williams,]:

$$K(x, x') = \theta_s^2 exp\left\{\frac{-||x - x'||^2}{2\theta_\ell^2}\right\} \tag{2.1}$$

where $\theta_t = \{\theta_s, \theta_\ell\}$, with θ_s and θ_ℓ being the hyperparameters of the covariance function regulating the y-scale and x-scale, respectively.

2.5 Results

Figure 2.2 presents the performance of various clustering strategies compared with generalized models using the predictive algorithms presented earlier. Before analyzing performance, we will present a brief interpretation of each grouping strategy. Using data from SMS, four groups were discovered as presented in Table 2.1. The group labeled *freq* are most actively engaged with text messaging on their phones, while *reg1* and *reg2* fall in the middle with *reg2* being more frequent than *reg1*. The most inactive group is labeled by *infreq*. In the profiles learned using the phone call logs, two groups were discovered: an active group and an inactive group in terms of their phone call level distributions. Notice that for the majority of time prior to EMAs, phone calls were rarely made by our study participants, and thus we see high percentages in the 'low' level. Using acceleration as a proxy to characterize participants' activity level, we found two: one *active* group and one *inactive* group. Again notice that the differences in the acceleration level distribution between the two learned groups are minor and only relative between them. With respect to locations, in the first group, the participants split most of their time between school and home; in the second group, the participants spent over 80% of their time at school at the expense of other places; and in the third group, the participants spent the majority of their time away from home (e.g., traveling out of town, visiting friends, and at leisure place of interests).

We also used cutoffs of 34 and 43 in SIAS scores to divide participants into low, medium, and high social anxiety groups [Heimberg et al., 1992]. In total, we experimented with 10 grouping approaches based on location, activity level, communications (SMS and phone calls), and SIAS scores as shown in Figure 2.2. Specifically, *DailyActivity* applies a combination of location, activity level, communications profiles; communication is based on the combination of phone calls and SMS (re-grouped into active and inactive) producing three groups (active in both SMS and calls, only active in either SMS or calls, inactive in both SMS and calls).

From Figure 2.2, using most of the grouping strategies, we were able to obtain better overall performance in lower weighted RMSE in our group models when compared

Table 2.1: Clustering based on communication, location, and acceleration data using G-means clustering algorithm.

				Group Profile (%)				
	Gp	Label	#Part	Low+	Low	Med	High	High+
SMS	1	reg1	22	80.5	16.8	2.0	0.5	0.2
	2	reg2	12	68.6	25.9	4.3	0.7	0.4
	3	infreq	9	93.7	5.9	0.3	0.1	0.0
	4	freq	19	49.1	36.1	8.6	3.5	2.7
Call	1	inactive	54		89.5	9.1	1.3	0.1
	2	active	8		65.7	29.2	4.4	0.7
Acc	0	active	25		83.1	4.6	12.3	
	1	inactive	37		91.2	2.7	6.1	
				Out	Edu	Friend	Home	Leisure
Loc	1	school-home	34	2.0	49.2	4.4	38.7	5.8
	2	school	18	3.0	83.0	2.7	4.9	6.5
	3	out	10	20.9	43.7	8.3	9.3	17.8

to the generalized model. Specifically, our generalized models using four different algorithms achieved a RMSE of 21.58 (random forest), 22.05 (Gaussian processes), 21.87 (linear lasso), and 22.31 (SVM), respectively. For each grouping strategy on Gaussian processes model, we were able to obtain average reductions of RMSE 0.8722 (Location), 0.6310 (activity level), 0.045 (SMS), 1.2330 (calls), 1.4505 (SIAS), 1.9268 (DailyActivity), 0.4264 (communication), 1.2675 (SIAS+communication), 2.1326 (All features - communi-cation), 1.9231 (All features - SIAS), respectively.

Note from Figure 2.2 that the DailyActivity grouping strategy consistently performed better than most other grouping strategies, and this strategy is also closest to the individual model approach (65 individual models for 65 participants) because it resulted in the most (25) subgroups among all these strategies, thus we used it to further investigate whether there are any specific patterns with respect to sample size to guide future design of group-level modeling approaches.

From Figure 2.3, we can see that there is a nonlinear relationship between

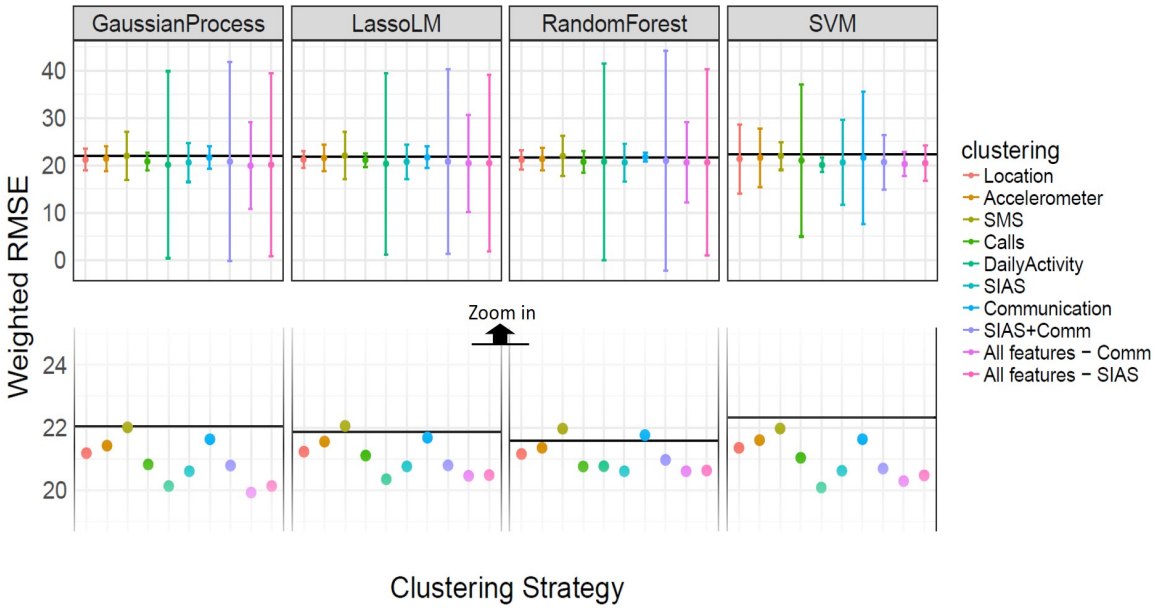


Figure 2.2: The performance of each grouping strategy compared with the generalized model's performance (black horizontal line). The y-axis is the weighted root mean square error (WRMSE). The error bars represent 2 standard deviations of each grouping strategy.

sample size of groups and their performances. Groups with small sample size tend to perform either extremely poorly or extremely well. This signals potential weak generalizability of profiling strategies that forms many small groups. So the ideal situation will be to form groups with profiling strategies that evenly distribute the samples across different subgroups.

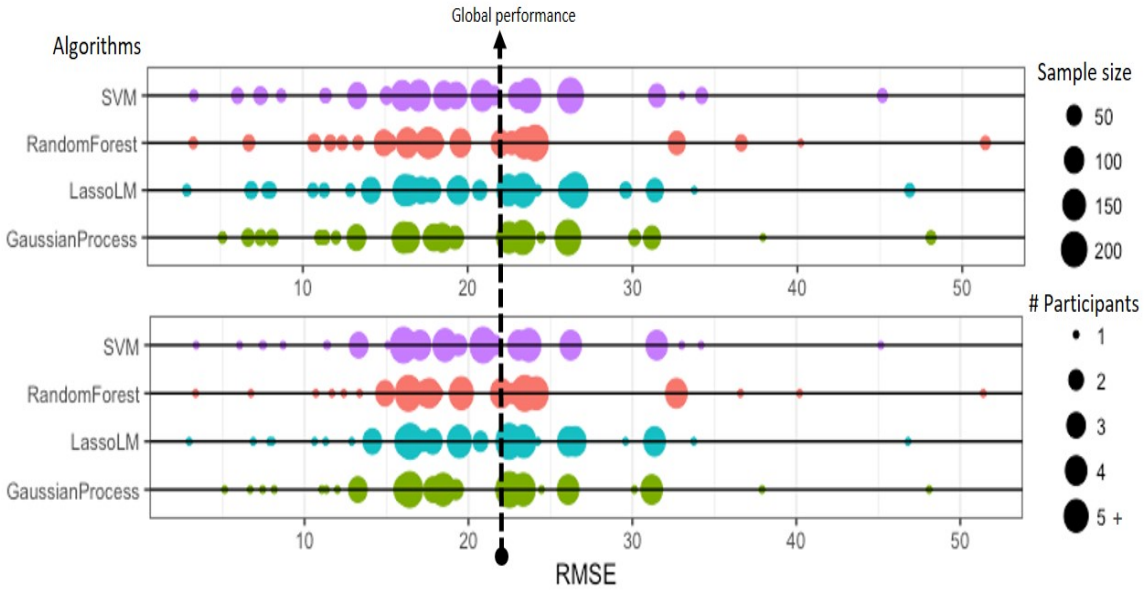


Figure 2.3: The impact of sample size on the performance of groups formed by DailyActivity strategy.

2.6 Summary

The focus of the present investigation was to provide a framework for accurately predicting negative affect from passively sensed data concurrent with individuals' affect ratings. Given that two weeks may be too short for algorithms to learn personalized models, we developed a method for predicting negative affect using a group-level approach. We first clustered participants using multimodal behavioral

profiling, then we predicted negative affect from passively sensed data. The results indicate that profiling users based on their behavior improves the performance of the predictive model compared to generalized models. Future work will study the predictability levels among the different groups using validated questionnaire measures of personality and depression. The present study contributes to a body of research that aims to use passively sensed data to recognize user affect and launch interventions when and where they are most needed.

3 | Offline Treatment Policy: An Emotion Regulation Case Study

Delivering treatment recommendations via pervasive electronic devices such as mobile phones has the potential to be a viable and scalable treatment medium for long-term health behavior management. But active experimentation of treatment options can be time-consuming, expensive and altogether unethical in some cases. There is a growing interest in methodological approaches that allow an experimenter to learn and evaluate the usefulness of a new treatment strategy before deployment. We present the first development of a treatment recommender system for emotion regulation using real-world historical mobile digital data from $n = 114$ high socially anxious participants to test the usefulness of new emotion regulation strategies. We explore a number of offline contextual bandits estimators for learning and propose a general framework for learning algorithms. Our experimentation shows that the proposed doubly robust offline learning algorithms performed significantly better than baseline approaches, suggesting that this type of recommender algorithm could improve emotion regulation. Given that emotion regulation is impaired across many mental illnesses and such a recommender algorithm could be scaled up easily, this approach holds potential to increase access to treatment for many people. We also share some insights that allow us to translate contextual bandit models to this complex real-world data, including which contextual features appear to be most important for predicting emotion regulation strategy effectiveness.

3.1 Introduction

Mental illnesses such as depression and social anxiety, if left untreated, can interfere with healthy life functioning, leading to lower disability-adjusted life years [Murray et al., 2013] and higher suicide rates [Bostwick and Pankratz, 2000]. It is estimated that more than 25% of Americans suffer from a diagnosable mental illness each year [Kessler et al., 2005b], yet half of them do not receive any treatment [America,] due to the scarce health care resources and limited access to traditional in-person care [Lin et al., 2018]. New mobile technologies and increasing smartphone ownership give rise to mobile health, a digital health care paradigm that creates opportunities to scale up health interventions to the underserved patient population [Hilty et al., 2013], especially those with chronic conditions.

One viable target for a digital health intervention that could benefit a significant portion of the population is emotion dysregulation, or difficulty selecting and effectively applying appropriate strategies to modulate the intensity or duration of emotional states [Gross, 1998]. Emotion dysregulation is observed broadly across many mental illnesses, and improvements in emotion regulation (ER) often accompany decreases in symptom severity [Fernandez et al., 2016,Sloan et al., 2017]. The ability to effectively manage negative emotions in our daily lives is of utmost importance. For example, days before a job interview, you may not be confident in your preparation, and feel anxious about it. You may find it difficult to focus on anything else, and cannot stop worrying about it or sleep. To manage your negative emotions, you might try a variety of strategies, including suppressing your thoughts about the upcoming interview, talking to a friend about it, conducting a mock interview for practice, distracting yourself with video games, or taking the advice from your therapist to identify and re-evaluate your catastrophic thoughts.

Ideally, one would conduct a randomized control trial (RCT) to evaluate the effect of different ER strategies in different contexts, but this can quickly become unfeasible if the intent is to evaluate more than a dozen strategies across different contexts. We address this challenge in part by using an offline contextual bandits to learn and evaluate a novel treatment recommender algorithm using an observational

dataset collected from a population of socially anxious individuals.

While an observational design necessarily limits what causal inferences are possible, our contributions in this work include the following: 1) to the best of our knowledge, this is the first study to apply Contextual multi-armed bandit (CMAB) on ER, a domain that is central to treatment for many mental illnesses; 2) we apply CMAB in an offline setting that learns an interpretable initial policy using observational data; 3) we leverage both passively (e.g., Accelerometer) and actively (e.g., how appropriate was timing of survey) sensed contexts with a designed reward signal using self-reported effectiveness to evaluate the CMAB performance using several different importance sampling based estimators, and compare them with both a random policy and the observed policy. Our results show significantly better performance in the proposed CMAB approaches in terms of the average reward of a policy, which we denote as usefulness.

3.2 Related Work

Emotion regulation (ER) has been studied in psychology for decades due to its importance in understanding how people manage their emotions [Gross, 1998], and its implications for both mental and physical health, and interpersonal relations [Aldao et al., 2010a]. People respond to stressful events using different ER strategies in different social and physical contexts, and according to different situational demands [Sheppes et al., 2014, Dixon-Gordon et al., 2015]. While ER strategies have long been considered as either adaptive or maladaptive, several researchers have argued that their effectiveness is context dependent [Bonanno and Burton, 2013, Aldao and Nolen-Hoeksema, 2013].

Notably, demographic characteristics such as age and gender [Nolen-Hoeksema and Aldao, 2011a], which may be considered internal contexts, significantly influence people's choice of ER strategies. In addition, numerous recent studies have focused on external contexts in people's daily lives, and investigated their impact on ER strategy choice [Troy et al., 2013, Aldao, 2013, Suri et al., 2018]. An ecological momentary

assessment study by Heiy et al. [Heiy and Cheavens, 2014] revealed that many of the most frequently used ER strategies were not the most effective for decreasing negative emotions, suggesting room for improvement in ER even among healthy individuals. To date, the capability of recommending the most effective ER strategies to people based on different contexts is urgently desired but remains a far-off goal [Doré et al., 2016]. In this work, we make an effort towards this goal to learn a personalized and adaptive approach for ER strategy recommendation across various contexts.

Many existing works propose various recommender systems targeting different health outcomes. For example, myBehavior, a mobile app that tracks user's physical and dietary habits, recommends personalized suggestions for a healthier lifestyle [Rabbi et al., 2015]. Cheung et al. [Cheung et al., 2018] created a mobile app called IntelliCare, which consists of a suite of 12 individual apps as 'treatments' that will be recommended for managing depression and anxiety. Yang et al. [Yang et al., 2018] created a mobile health recommender system that integrates depression prediction and personalized therapy solutions to patients with emotional distress. In their system, personalization is realized using 9 external factors related to depression, including family life, external competition, interpersonal relationship, self-promotion burden, economic burden, work pressure, individual personality, coping style, and social support, which are assessed using mobile questionnaires. These mobile health efforts are consistent with a mobile intervention framework called Just-in-time adaptive intervention (JITAI) [Nahum-Shani et al., 2017].

Two aspects regarding the intervention decisions made in a JITAI framework are the timing of intervention delivery and choosing the best intervention strategy to deliver. Most existing works focus on optimizing for the best timing to deliver an intervention (e.g., predicting stressful moments linked to emotional eating [Rahman et al., 2016]). By contrast, our work focuses on identifying the most effective ER strategies based on a person's context. Reinforcement learning with Markov decision processes (MDPs) are typically used to operationalize the key objectives of a JITAI. Example applications include personalizing sepsis treatment strategies [Peng et al., 2018], encouraging physical activity for diabetes patients [Yom-Tov et al., 2017], and managing stress [Jaimes et al., 2014]. Interestingly, although reinforcement learning

is not directly applied to recommend ER strategies for emotion management, it has been applied to understand the psychological and cognitive process of ER [Marinier et al., 2008, Raio et al., 2016].

In this work, we propose to leverage contextual multi-armed bandits, a reinforce -ment learning algorithm that treats each learning sample as independent from the same underlying data generating the distribution, but ignores the long term impacts on the distal outcome [Dudik et al., 2011]. CMAB has been mainly applied in domains such as web contents and advertisement placement [Li et al., 2010, Tewari and Murphy, 2017]. In recent years, it has also been applied in numerous mobile health applications, such as hospital and doctor referral for medical diagnosis [Tekin et al., 2014], personalized feedback for healthier lifestyle [Rabbi et al., 2017], and physical activity recommendation [Liao et al., 2020a]. Unlike these studies, which were conducted in an online setting or with simulations, our work focuses on the off-policy setting, in which a historical dataset on ER from a mobile health study is used to train an initial warm-start recommendation policy on ER. We design the various reinforcement learning components in the context of recommending ER strategies, and applied various importance sampling based techniques in learning and evaluation.

3.3 Contextual Multi-armed Bandit for Emotion Regulation

Contextual multi-armed bandit (CMAB) is an reinforcement learning algorithm that leverages contextual information to learn a policy that triggers actions based on the context to achieve optimal expected rewards. Typically, CMAB consists of an agent that interacts with an environment over a finite number of trials $i = 1, 2, \ldots, T$ such that: 1) it observes a context x from an input space $\mathbf{X}$; 2) chooses an action from a set $\mathbf{A} = \{a_1, a_2, \ldots, a_{k-1}, a_k\}$, which contains all the strategies that each corresponds to an arm of a k MAB; and 3) receives a reward signal r_i. The goal of the agent is to learn a policy to guide action decisions. Unlike a full-blown reinforcement learning algorithm typically modeled using MDPs, where an action decision impacts future

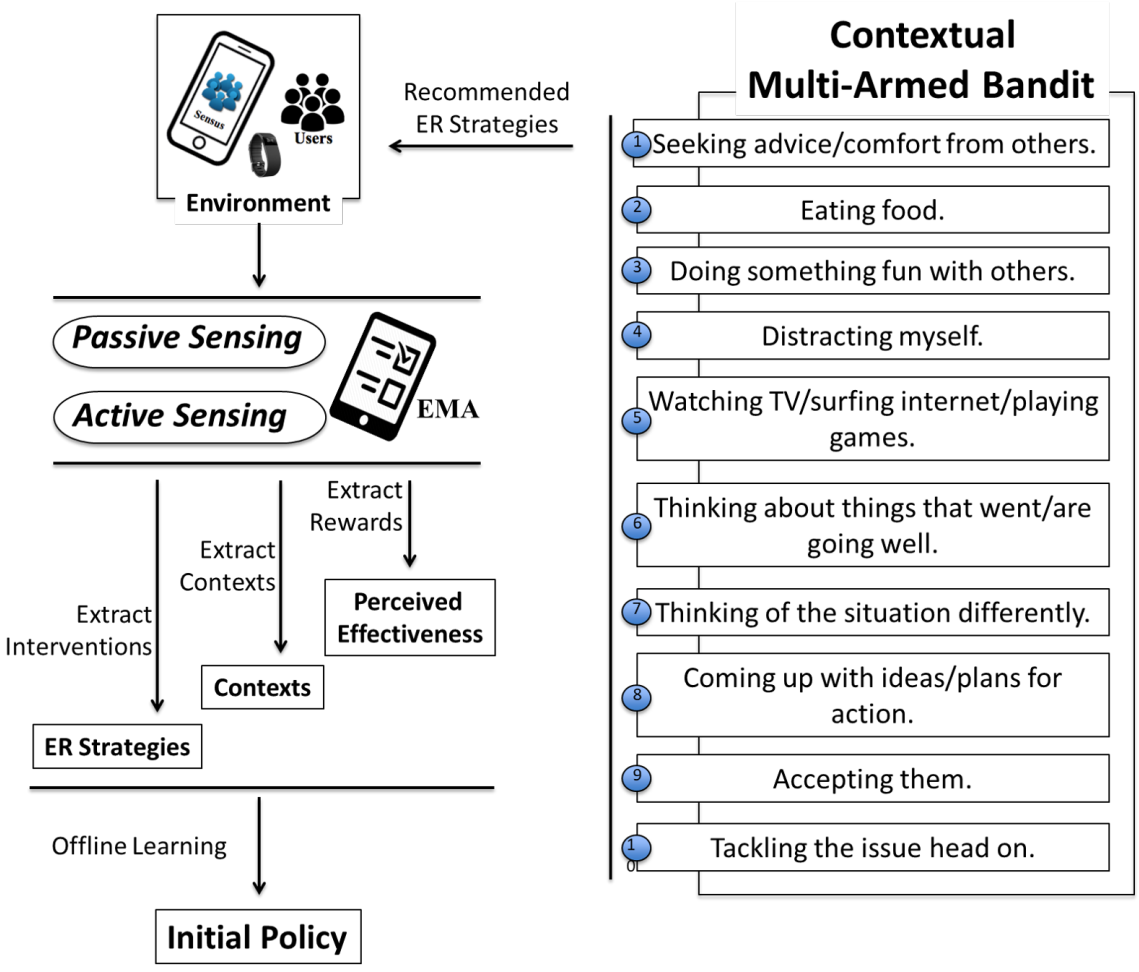


Figure 3.1: Learning initial policy for emotion regulation (ER) using offline learning in contextual multi-armed bandit.

states and action selections, CMAB assumes that $\{(x_i, a_i, r_i)\}_{i=1}^{T}$ are independently and identically distributed following an unknown generative distribution $\mathbf{D}$.

We formulate ER recommendation as a CMAB using mobile sensing technologies as shown in Figure 3.1. Smartphones and wearables are applied to track the users both passively with sensor embedded devices and actively with mobile ecological momentary assessments (EMAs). These mobile sensing data streams will be processed into the contexts, the recommended ER strategies, and the associated rewards for our CMAB framework.

In the offline learning, observational data generated under a different policy will be used to learn and evaluate an initial policy. This data-generating policy is called the behavior policy and can be denoted as π_b. Similarly, the initial policy is called the target policy denoted as π_e.

We seek to achieve two objectives: 1) Learn an initial policy π_e^* given an observational dataset, called the learning problem which is formulated as

$$\pi_e^* = \underset{\pi_e \in \Pi}{\operatorname{argmax}} \mathbf{V}^{\Pi}. \tag{3.1}$$

Where $\mathbf{V}$ represents the value of a policy and Π, the function class of possible policies. 2) Evaluate the performance of the initial policy using expected rewards from the testing samples. We call this the evaluation problem and this is formulated as

$$\mathbf{V}^{\pi_e} = \mathbf{E}_{(x,r)\sim D}[r_{\pi_e}(x)]. \tag{3.2}$$

In the next section, we present the technical details on both the learning and evaluation problem to learn and evaluate the initial policy.

3.4 Learning and Evaluation in Contextual Multi-armed Bandit

We consider learning in a linear policy class, of which the candidate policies are efficient for learning and easy to interpret. We apply importance sampling techniques that use a certain form of weighting scheme denoted as $\frac{\pi_e(a_i|x_i)}{\hat{\pi}_b(a_i|x_i)}$ in context x_i to correct for the distributional shift between the target and behavior policy in order to have an unbiased estimate of the target policy value [Dudík et al., 2014].

There are three main value estimators that lie at the core of offline policy learning and evaluation within the contextual bandit framework; namely, the direct method (DM), Inverse Propensity Weighting (IPW), and Doubly-Robust (DR) [Atan et al., 2018]. None of these approaches are guaranteed to perform optimally in every application scenario. Thus, we apply all of them in learning the optimal policy, and report their results. Below, we provide more details on the benefits and drawbacks of each approach.

The Direct Method (DM)

The direct method, sometimes called the response surface modeling or covariate adjustment, is the family of approaches that consist of learning a predictive model which maps context and actions to the rewards in a regression model. Specifically, the direct method (DM) consists of estimating a reward approximator for $\hat{r}(x, a)$, where $\hat{r} : X \times A \rightarrow R$. This will result in a value function:

$$V_{DM} = \frac{1}{T} \sum_{i=1}^{T} \pi_e(a_i|x_i)\hat{r}(x_i, a_i) \tag{3.3}$$

where, π_e is the target policy. While this approach is simple to implement and can be used with most regression models, it relies heavily on model specification and overlap in the distributions of the behavior and evaluation policies. This gets even more complex in application domains where the physical process of the underlying environment is not well understood. In effect, most of the direct methods approaches

suffer from high bias in the estimates, albeit with low variance for a sufficiently well-specified model. Some popular examples of algorithms using this approach for learning counterfactual predictions are the Bayesian Additive Regression Trees (BART) [Chipman et al., 2010] and the Causal Forest [Wager and Athey, 2018].

The Inverse Propensity Weighting Method (IPW)

The inverse propensity weighting approach seeks to correct for the distributional shift caused by the behavior policy by using the behavior policy π_b if known or an estimate $\hat{\pi}_b$ (also known as propensity scores [Chakraborty, 2013]) otherwise. The correction in distribution shifts is achieved using importance sampling in the estimator to evaluate the target policy. Mathematically, a generalized IPW estimator called the trimmed IPW (tIPW) is as follows:

$$V_{tIPW} = \frac{1}{T}\sum_{i=1}^{T}\frac{\pi_e(a_i|x_i)}{max\{\hat{\pi}_b(a_i|x_i), \tau\}}r_i. \tag{3.4}$$

Where τ is a lower bound on the propensity scores to reduce the effect of large weights on variance of the estimator. When $\tau = 0$ this reduces to a classic IPW estimator. When $\tau = 0$ this approach gives an unbiased estimate of the value of the target policy, however it suffers from high variance due to extreme values of propensity scores (e.g., a propensity score close to zero will give rise to approximately infinite weights). Some examples of algorithms using this approach are the Policy Optimizer for Exponential Models (POEM) [Swaminathan and Joachims, 2015] and the Offset tree [Beygelzimer and Langford, 2009].

The Doubly Robust Estimator (DR)

The doubly robust approach combines the DM and IPW methods to achieve a balanced trade-off between bias and variance. This avoids extremely high bias and variance in the estimator. The DR estimator has been formalized by Dudík et

al [Dudík et al., 2014] as follows:

$$V_{DR} = \frac{1}{T}\sum_{i=1}^{T}\left[\hat{r}(x_i, a_i) + \frac{\pi_e(a_i|x_i)}{\hat{\pi}_b(a_i|x_i)}\left(r_i - \hat{r}(x_i, a_i)\right)\right]. \tag{3.5}$$

The DR estimator combines the DM (typically a maximum likelihood estimator) with the importance sampling of the residual from the DM approximator. This is described as doubly robust because if the DM model is correct, then the expected residual from the model $E_Y[\hat{\varepsilon}] = 0$, leaving the second term equal to zero for any arbitrary behavior policy $\hat{\pi}_b$; similarly, if the $\hat{\pi}_b$ is correctly estimated, then the second term is a consistent estimator of the error bias from the DM approximator. Though more robust, DR is error prone when both the DM and the behavior policy approximators are misspecified [Kang et al., 2007].

Propensity Score Estimation

As noted above, the behavior policy that generated the data is unknown and needs to be estimated from the data. This is achieved by estimating propensity scores, which represent the likelihood of choosing strategies in different contexts. Propensity scores also serve to reduce multivariate contextual data [Rosenbaum and Rubin, 1983] into one-dimensional scores such that treatment group distributions are matched. The goal of the propensity scores is to create a pseudo-population where the effect of selection bias due to unobserved confounders, as evidenced by distributional mismatch across strategies, is minimized.

Ensuring overlap in the strategies with respect to the propensity scores reduces the possibility of extreme values in the IPW and DR estimation, given these approaches depend on the estimated score denoted $\hat{\pi}_b(a|x)$ or $\hat{P}_{ij}$ in the algorithm 1. Estimation methods such as logistic regression have typically been used but they are limited due to their linearity assumption [McCaffrey et al., 2013]. Recently, there are non-parametric machine learning models developed to add more flexibility in order to model more complex data, such as what we usually expect in human data. An

Algorithm 1 Generalized Algorithm for Policy Learning

Input: $\mathbf{X}$, $\mathbf{A}$, $\mathbf{R}$.
Output: $\pi^*(\mathbf{x})$.

// **Propensity Score Estimation**
Fit Generalized Boosted Model $\hat{f} : X \to A$ on $S_N = (X, A)$ to balance covariate distribution.
Obtain propensity score matrix $\hat{P} = \hat{f}(x)$.
// **Reward Imputation**
fit a one-time logistic regression $\hat{r} : X \times A \to R$ for each strategy
for $r_{ij} \in \mathbf{R}$ (a matrix of rewards). **do**
    if DM method **then**
        $\hat{r}_{ij}^{DM} = \hat{r}(x_{ij}, a_{ij})$
    end if
    if IPW method **then**
        $\hat{r}_{ij}^{IPW} = \frac{r_{ij}}{\hat{P}_{ij}}$
    end if
    if DR method **then**
        $\hat{r}_{ij}^{DR} = \hat{r}_{ij}^{DM} + \frac{(r_{ij} - \hat{r}_{DM})}{\hat{P}_{ij}}$
    end if
end for
Set $\hat{R} = \{\hat{r}_{ij}\}_{i=1:T, j=1:k}$ the weighted reward matrix
// **Policy Optimization**
Fit logistic regression $\hat{h} : x \to \hat{R}$ on new training set (X, R).
// **For policy** $\pi^*(x)$
$\pi^*(x) = \text{argmax}_{a \in A} \hat{h}(r_a|x)$
return $\pi^*(x)$.

example of a non-parametric model is Generalized Boosted Models (GBM). GBM estimation uses an iterative process with multiple regression trees to capture nonlinear relationships between strategies and context variables without over-fitting the data. We implemented GBM propensity score estimation in our analysis using the R package *twang* [Ridgeway et al.,]. We used the absolute standardized mean difference [Stuart et al., 2013] as the stopping criteria over 5000 iterations.

The Learning Algorithms

In our experiments we used a multivariate logistic regression as the value function approximator that maps contexts to rewards for each ER strategy within the direct method and doubly robust estimators. We used logistic regression with ℓ_2 regulariza -tion for the ease of interpretation and replication in other studies. We will call the learner using direct method (DM) and the one using doubly-robust estimation as (DR) in our experimentation. The offset tree, denoted OT, is different in that it learns several binary regression trees for propensity weighted reward in each offset tree. More details can be found in Beygelzimer et al. [Beygelzimer and Langford, 2009]. We compare the performance of these three approaches, and benchmark them against a random policy (i.e., randomly choosing one strategy from the 10 ER strategies) and the observed policy (i.e., what people reported using in the data).

The Evaluation of Learned Policies

Given the selection bias in the test data, we evaluate the performance of the different recommender algorithms using two variants of importance sampling approaches; name -ly, the inverse propensity weighting (IPW)(e.i.$\tau = 0$) and the trimmed inverse propensity weighting (tIPW)(e.i. $\tau \neq 0$) (see equation 4.5) by varying the parameter τ.

We consider both approaches because while the IPW provides an unbiased estimate of the mean policy reward with possibly high variance, the tIPW reduces the variance at the cost of more bias in the estimator. τ is a nuisance parameter and can be determined heuristically if $\tau < 1/k$, where k is the number of strategies according

to lemma 3.1 of Strehl et. al, [Strehl et al., 2010]. We compare the performance of each algorithm on the average reward on the test set.

3.4.1 Design of ER Recommender System

Our contextual variables capture the user's state around the time to use a strategy. They are summarized in Table 3.1. A combination of these variables allows us to provide contextual recommendation for ER strategies. For example, given that a user is at home in the evening with a trait social anxiety level of 30, we would recommend tackling issues head on if our algorithm predicts it to be the most effective strategy. The actions in our formulation are the top 10 most frequently used adaptive strategies, which are shown in the CMAB in Figure 3.1. Admittedly, there are multiple ways to reduce dimensionality of the ER feature space and we explored additional approaches in other analyses. However, we chose to focus on this subset of strategies as they are mostly considered healthy strategies (i.e., they tend to be associated with positive health consequences, unlike a strategy such as using alcohol or drugs to change one's feelings) and were most frequently reported in our learning data.

The reward signal needs to reflect the effectiveness of the chosen strategy in the given context at helping to manage the participant's emotion. In our data, participants reported the perceived effectiveness of their ER attempt on a scale of 0-10. We binarized this outcome measure to define a reward signal for the agent. Our threshold was defined as the average of effectiveness scores across all users, or the grand mean. Let $O(x_i, a_i)$ denote the immediate effectiveness of the chosen ER strategy at time i in context x_i, we have the grand mean as

$$\hat{O} = \frac{1}{N} \sum_{i=1}^{T} O(x_i, a_i), \tag{3.6}$$

. The reward signal for each context x and action a is thus defined as:

$$r(x, a) = \mathbf{1}_{\{O(x,a) > \hat{O}\}}, \tag{3.7}$$

where **1** is an indicator function that returns 1 when the condition is satisfied, and 0 otherwise.

Table 3.1: Contexts for the proposed contextual multi-armed bandit algorithm.

Context	Description
Social partners	self-reported social relationship with people in the context (e.g., being with classmates, friends, strangers/acquaintances, romantic partner and family).
Social interaction	self-reported social context (e.g., being alone, no interactions with others or being around them, and interaction with others).
Social preference	self-reported social preference (e.g., more people, less people).
Motivation to change	self-reported motivation to change feelings on a 0-10 scale.
Device OS	device platform (e.g., Android and iOS).
Social anxiety score	self-reported social anxiety score using SIAS scale with 0-80 range.
Time of day	this is a manual binning of periods of time in the day, (e.g., morning, mid-day, afternoon, and night).
Semantic Location	Self-reported locations (e.g., the gym, home, in transition between locations, other homes, other locations, religious places, restaurant, school, shopping center or workplace).
Accelerometer	passively sensed measure of user movement (e.g., mean, energy and standard deviation).
Activity Type	passively recognized human activity types (e.g., cycling, stationary, walking and automotive).
Appropriateness of Timing	self-reported measure of how appropriate the timing was for sending an survey prompt on a scale of 0-10.

3.5 Experiments

3.5.1 Study Design

After getting approval from the university's Institutional Review Board (IRB), $N = 114$ participants aged 18 years and older were recruited in a US college department and community to enroll in the present study. Participants were eligible to enroll if they scored at least 29 on the Social Interaction Anxiety Scale (SIAS) developed by Mattick & Clarke [Mattick and Clarke, 1998]. This cutoff was selected to recruit a sample experiencing moderate to severe social anxiety symptoms (scale range is 0-80). Four participants were excluded in the analysis due to missing data; specifically, 1 participant did not report any EMA data and 3 participants did not have any reports of effectiveness of an ER strategy, leaving 110 participants with the following demographics: 81 female, 29 male (no participants reported a non-binary gender identity); 86 undergraduates, 11 graduates or professional students, and 13 others; aged 18-34 with mean 20.41 and SD 2.98; 82 reported their race/ethnicity as White, 21 Asian, 7 African American, 3 Middle Eastern, 3 Native Hawaiian/Pacific Islander (numbers add up to more than 110 because some participants identified as multiple races). Their SIAS scores ranged from 29 to 73 ($M = 46.68, SD = 10.39$). Although the full SIAS was used for recruitment (for comparison to the reference group from prior published work), the sum of the straightforwardly-worded items was used for analyses, because the straightforwardly-worded items have been shown to have prefer-able psychometric properties to the full scale [Rodebaugh et al., 2007].

A mobile app called MetricWire was installed on all participants' personal smartphones to collect random time survey data for five weeks. Six identical surveys were sent randomly within each two hour window from 9am to 9pm daily. Participants were instructed to complete the surveys as promptly as possible upon receiving the notifications. If participants had not completed the survey within 30 minutes of the initial notification, the app sent a reminder notification. If not completed after 45 minutes, the survey disappeared. Participants were instructed to answer the survey with reference to when they received the initial survey notification. This instruction

might introduce a small degree of recall bias into survey responses, but was included to enhance ecological validity by sampling a wide variety of situations in daily life, including situations in which it would be difficult to respond to a survey immediately (e.g., when a participant is taking an exam or in the middle of a conversation). Sensor data were also passively collected from participants' smartphones to capture their activity levels and GPS location. Table 3.1 summarizes the contextual features extracted from both survey and passive data.

3.5.2 Data Processing

We used both the random time survey data and the sensor data from the study to obtain the contexts surrounding the reported ER strategy use and its effectiveness. All contextual variables are aligned with random time prompts using two hour windows. For example, accelerometer data within two hours prior to each survey starting time were aggregated to capture the level of activity for each reported ER strategy use. We transformed the x, y, z dimensions of the accelerometer using the formula $\frac{1}{3}\sqrt{x^2 + y^2 + z^2}$ to obtain an orientation invariant measure for acceleration. Activity type data consisted of the activities recognized by MetricWire. These activities include stationary, walking, running, automotive, and cycling. The feature associated with each activity type is the sum total of its occurrence in the two hour window. Semantic locations such as home, transition, religious place, restaurant, school, shop -ping, someone else's house, work etc., provided by participants in the surveys were included in the context variables. Temporal features were created using four time windows: morning (9-12PM), mid-day (12 -3PM), late-afternoon (3-6PM), and night (6-9PM). Finally, we included other survey responses, such as rating the convenience of responding to the prompt when fired, and others summarized in Table 3.1 as context variables.

The original EMA data consists of 12742 learning samples from all participants. We excluded samples where participants did not report an effectiveness score for using ER strategies, either because they reported that they did not try to change their feelings (which is one option in the menu provided; 7617 samples were excluded for

this reason) or because they used a strategy but skipped the survey prompt about effectiveness (239 samples were excluded for this reason). This leaves 4886 learning samples. 259 samples where important survey responses were missing (specifically, any missingness on reported convenience of responding to the prompt when fired, semantic location, or motivation to change feelings) were further excluded, leaving 4627 samples for analysis. We avoided imputing the 259 samples as these are self-reported ground truth data. On the other hand, we used multiple chained imputation to impute data on the passively sensed accelerometer and activity type data, which have missing rates of 65% and 68%, respectively. The MICE R package with classifica -tion and regression trees method was used for the imputation.

The remaining 4627 learning samples consisted of instances where participants reported choosing not only one strategy but also combinations of strategies in a menu of 20 strategies available to them in the survey. Our algorithm, however, considers the effect of a single strategy at a time. To accommodate this constraint, we split the samples in which more than one strategy was reported to have been used into multiple independent samples. For example, if a participant used a combination of eating food and distracting themselves in a given context, we treated this case as two separate samples in which a single strategy was used, and assigned the same effectiveness score to both. This allows us to retain all the data in which effectiveness was reported, increasing power, and not cut the common occurrence in which people report using more than one ER strategy, increasing generalizability. While we recognize this may reduce accuracy in parameter estimation as more bias is being introduced with this data augmentation approach because it is possible that the self-reported effectiveness score does not apply to all applied strategies equally, we felt the benefits for data retention and external validity were worth the trade-off. By augmenting the data this way, we obtain 6259 learning samples, including instances where any of the top 10 most adaptive strategies have been used. By contrast, restricting the data sample to instances where only one strategy was reported being used by the participants, we ended up with 2496 samples, which is about 1/3 of the data generated by the augmentation approach (contact the first author to see results for the CMAB analyses using this smaller dataset).

We used a total of 40 contextual variables summarized in Table 3.3, consisting of binary variables (e.g., semantic locations, social partner(s) vs. alone, etc.) and continuous variables, including convenience of responding to the prompt when fired, motivation to change, SIAS score (SIASsf), activity types, and accelerometer features. The continuous variables were scaled to a range between $[0, 1]$ to avoid biasing coefficient estimations toward the continuous variables.

3.6 Results

Our results as summarized in Table 3.2 show the mean reward across the different recommender algorithms and baselines. We report the mean reward with standard errors on a 5-fold cross validation (due to relatively small data), and test for level of significance using an independent samples t-test at $\alpha = 0.05$. The parameter τ regulates the effect of extremely large weights due to low propensity scores by capping all scores below the chosen value of τ. Note also that τ uses the same value in both learning and evaluation for each policy. The algorithm learned with doubly robust estimator (DR) outperforms all its competitors, including the Offset Tree (OT) and the DM learner. This can be seen from its absolute mean reward and the tight confidence bounds for all values of τ. This implies that the doubly robust method achieves the right trade-off between high variance and high bias, at least relative to the other approaches tested, making it a more more reliable statistical estimator of off-policy performance in our data. We also see that the gap between the Offset tree and the DR get closer as the value of τ is increased. This is as expected as the classic OT algorithm is heavily dependent on the inverse propensity weighting and thus more affected by high variance. Notice that the parameter values of τ are set below 0.1 to match with the theoretical constraint developed in Lemma 3.1 of [Strehl et al., 2010]. Also note that the DM, Random, and Observed policies are affected by the parameter τ only in the evaluation stage, but they still benefit from less variance in the mean reward estimation on the test set.

Table 3.2: Mean reward by policy (mean ± std). Superscripts † and ∗ respectively represent statistical significant at $\alpha = 0.05$ over random and behavior policy baselines.

Algorithm	IPW($\tau = 0$)	IPW($\tau = 0.02$)	IPW($\tau = 0.05$)
DR	$11.48 \pm 2.07^{\dagger *}$	$10.84 \pm 1.96^{\dagger *}$	9.11 ± 1.27
DM	11.08 ± 2.38	10.46 ± 2.26	8.65 ± 1.28
OT	8.60 ± 1.70	8.41 ± 1.60	7.81 ± 1.42
Observed	8.25 ± 0.40	8.20 ± 0.40	7.84 ± 0.32
Random	8.24 ± 0.51	8.20 ± 0.50	7.91 ± 0.49

To probe deeper into a qualitative evaluation of the DR algorithm, we examine the effect sizes of several contextual variables in the learning stage in terms of how they predict individual strategies. These effect sizes are summarized in Table 3.3. Contextual variables with a positive effect size can be interpreted as increasing the odds of positive rewards if that strategy is chosen within that context and vice versa for negative effect sizes. For example, the chances are high the strategy will be perceived as effective if the user is recommended to seek advice or comfort from others when they have recently been stationary because the effect size is 1.07. Note that the effect sizes in bold are statistically significant at $\alpha = 0.05$.

While there are many significant effects, pointing to the importance of many contextual factors in ER, a few context variables are notable for their large effect sizes. Overall, the contextual predictors that tended to have the largest absolute effect sizes (indicating that they are the most important in determining effectiveness) are the convenience of responding to the prompt when fired, motivation to change thoughts/feelings, trait social anxiety symptoms, accelerometer features, and certain activity types (see Figure 3.2 for a ranking of contextual features from most to least important, as defined by the absolute value of their effect sizes). This suggests that a person's movement helps to determine what ER strategies are most likely to help them feel they have effectively regulated their emotions. The predicted effectiveness of strategies increased when it was a convenient time to be interrupted with a survey prompt, pointing to the importance of timing in interventions (and suggesting that JITAIs may be a step in the right direction). Notably, effect sizes for time of day were smaller than effect sizes for convenient time for interruption, suggesting personalized timing for ER strategy implementation may be particularly helpful. Strategies were

Strategies	Social Partners					Semantic Locations									
	Classmates	Friend	Strangers	Romantic	Family	Gym	Home	Transit	Other Home	Other Loc	Religious Loc	Restaurant	School	Shopping	Work
S1	**-0.18**	0.03	**0.12**	-0.03	0	**0.19**	0.01	**-0.1**	0.03	-0.05	0.04	**0.06**	-0.04	**-0.13**	-0.03
S2	**0.1**	**0.16**	0	0.05	**0.03**	0.05	0.03	**0.2**	**-0.18**	**-0.07**	**-0.41**	**0.11**	**0.14**	0	**0.13**
S3	**0.05**	**-0.23**	**-0.14**	**-0.25**	**-0.36**	**-0.19**	**0.06**	0.03	0	**0.18**	-0.01	**0.09**	0.01	-0.01	**-0.15**
S4	**0.08**	**0.05**	**0.05**	**-0.08**	0.05	**0.08**	**-0.05**	0.02	0	-0.03	**0.01**	-0.02	**0.03**	0	**-0.05**
S5	0.02	0	**0.24**	-0.01	-0.06	**-0.44**	**0.06**	**0.06**	**-0.04**	**-0.17**	**0.04**	**0.13**	0.05	**0.18**	**0.13**
S6	-0.01	**0.12**	**0.07**	-0.03	-0.01	**0.2**	-0.01	0.03	**0.11**	**0.09**	**-0.54**	**-0.19**	**0.07**	**0.16**	0.06
S7	**-0.07**	**0.08**	0.02	**0.05**	**0.07**	**0.23**	**-0.07**	-0.03	**-0.19**	**0.07**	**-0.06**	**-0.2**	**0.05**	**0.19**	0.01
S8	0.01	**0.04**	**0.1**	-0.04	-0.02	**-0.13**	-0.01	**-0.04**	**-0.06**	0.01	**0.26**	**-0.09**	**0.06**	0.06	**-0.07**
S9	**0.15**	**0.14**	**0.06**	**0.1**	**0.07**	**-0.08**	**-0.12**	0.04	**-0.1**	-0.01	**0.39**	0.01	**-0.15**	**0.22**	**-0.21**
S10	0.01	**0.08**	**0.05**	-0.04	**0.17**	0.06	**-0.05**	**0.1**	-0.02	0.01	**-0.14**	0.02	0.02	0.02	-0.02

	Other EMA			Time of Day				Activity Types				
Strategies	Appropriate	Motiv2Change	SIASsf	Morning	Mid-Day	Afternoon	Night	Stationary	Walking	Running	Automotive	Cycling
S1	**1.07**	-0.11	-0.12	-0.01	**0.06**	**0.03**	**-0.08**	**1.07**	**-1.3**	**0.35**	**0.67**	**0.27**
S2	-0.08	**-0.16**	**-0.37**	0	**-0.01**	**-0.05**	**0.06**	**0.26**	**0.89**	**-0.07**	**0.62**	-0.05
S3	-0.12	**0.85**	-0.13	**-0.12**	0.03	**0.05**	0.03	0.33	**0.81**	**-0.09**	-0.06	**-0.07**
S4	**0.69**	**-1.11**	**-0.51**	**-0.03**	**0.04**	**-0.01**	0	**0.71**	0.1	**-0.04**	**0.6**	**-0.05**
S5	0.29	-0.02	-0.09	0.01	0	0.01	**-0.01**	**-0.15**	-0.1	0	**0.37**	0.03
S6	**0.74**	**-0.39**	**-0.29**	**-0.1**	0.02	0.01	**0.07**	**0.28**	**0.23**	**0.35**	**0.39**	**-0.19**
S7	**1.17**	**-0.44**	**-0.55**	0.01	**-0.02**	**-0.02**	**0.04**	-0.15	**-0.74**	0	**0.53**	**-0.07**
S8	**0.7**	**-0.29**	**-0.85**	**-0.07**	**0.01**	0	**0.06**	**0.5**	0.01	**0.53**	**0.52**	**-0.49**
S9	**0.84**	0.21	**-0.47**	**-0.03**	**0.03**	-0.01	0.01	0.21	**-0.5**	**0.64**	**0.89**	**0.37**
S10	**0.84**	-0.12	**-0.35**	**0.03**	-0.01	0	**-0.02**	**-0.61**	-0.06	**-0.24**	**0.65**	**-0.6**

Strategies	Accelerometer			Platforms		Social Interactions			Social Preference				
	Mean Acc	Std Acc	Energy Acc	Android	iOS	Alone	Interacting	Around	A lot Fewer	Slightly Fewer	Same	More	A Bit More
S1	0.29	**-1.4**	**-0.67**	0.03	-0.03	**-0.09**	**0.13**	**-0.04**	**-0.1**	**-0.14**	**0.08**	**0.14**	0.01
S2	-0.21	**-0.84**	**-1.3**	**0.06**	**-0.06**	-0.05	0.02	0.03	**-0.05**	**-0.13**	**0.13**	**-0.14**	**0.19**
S3	**-0.61**	0.1	-0.11	**0.11**	**-0.11**	**-0.18**	**0.1**	**0.08**	-0.07	**0.04**	**0.21**	**-0.21**	0.04
S4	**0.84**	**-0.64**	0.13	0.01	-0.01	-0.01	**0.08**	**-0.07**	**-0.16**	**-0.08**	**0.08**	**0.09**	**0.08**
S5	**0.51**	**-1.13**	**-0.81**	**0.04**	**-0.04**	**-0.09**	**0.13**	**-0.04**	**-0.17**	**-0.14**	**0.06**	**0.14**	**0.1**
S6	**0.59**	**0.45**	-0.11	0.02	-0.02	**0.04**	0	**-0.04**	**0.06**	-0.04	**0.09**	**-0.15**	**0.05**
S7	**0.35**	**-0.24**	0.22	**0.02**	**-0.02**	**-0.07**	**0.09**	-0.01	0.01	**-0.07**	**0.1**	**-0.15**	**0.11**
S8	0.18	**-0.47**	**0.29**	**0.05**	**-0.05**	-0.01	**0.04**	**-0.03**	**-0.08**	-0.03	**0.11**	**-0.1**	**0.1**
S9	0.07	**0.34**	**1.45**	0	0	**0.02**	**0.02**	**-0.03**	-0.02	**-0.06**	**0.11**	0.02	**-0.04**
S10	**-0.89**	0.03	**-0.68**	-0.01	0.01	0.01	**0.04**	**-0.05**	**-0.09**	-0.01	**0.05**	-0.03	**0.08**

Table 3.3: Coefficients(*rounded to 2 decimal places*) of Contextual Predictors of Strategies (Strats). The strategies are mapped as follows; *Seeking advice/comfort from others*(S1), *Eating food*(S2), *Doing something fun with others*(S3), *Distracting myself*(S4), *TV/internet/gaming*(S5), *Thinking about things that went/are going well*(S6), *Thinking of the situation differently*(S7), *Coming up with ideas/plans for action*(S8), *Accepting them*(S9) and *Tackling the issue head on*(S10)

predicted to be less effective for more (vs. less) socially anxious participants, even among this sample where all participants were elevated in social anxiety symptoms at baseline), providing further evidence of emotion dysregulation in this population. Higher motivation to change thoughts/feelings predicted higher effectiveness ratings tied to the ER strategy 'doing something fun with others,' but lower effectiveness ratings tied to the ER strategy distraction, demonstrating that contexts can change the effectiveness of different strategies in opposing directions.

3.7 Discussion

This study provides evidence that a contextual bandits recommender algorithm may be used to improve ER, based on the current finding that the best performing algorithm, the learner with doubly robust estimation (DR), outperforms the observed ER of socially anxious participants. Further, contexts matter for effective ER, based on our finding that the DR algorithm also outperforms the random algorithm.

The results from this chapter have broad implications for the design and analysis of future recommender systems algorithms. By leveraging the abundance of available observational data from previous studies or interactive systems, a researcher might be able to estimate the usefulness of a novel recommender algorithm before deployment. Recent theoretical studies [Zhang and Bareinboim, 2017] suggest that combining offline policy learning together with online approaches leads to data efficient exploration and adaptations in the online setting. This could potentially reduce the user attrition or disengagement problem that plagues most interactive systems and ecological moment -ary assessment studies [Tewari and Murphy, 2017]. In addition, a researcher could use this method to determine the most critical features that affect the effectiveness of ER strategies in order to collect the most salient data for a new study when resources are limited.

Some of the strategies included in this recommender algorithm are cognitive, meaning that they involve a change in thinking (e.g., accepting thoughts/feelings), whereas others are behavioral, meaning that they involve a change in actions (e.g.,

eating food). Notably, contexts do not seem to have the same effect on strategies of the same cognitive/behavioral type. For example, our findings indicate that walking makes it more likely that thinking about things that went/are going well will be an effective strategy, and less likely that thinking of the situation differently will be effective. The distinction between these two specific strategies is subtle; for one, you are trying to think of positive things that may or may not be related to the situation at hand, and for the other, you are focused on the situation at hand but trying to notice other aspects of it or conceptualize it in a different way.

Regarding the social strategies in this recommender algorithm (those that use other people to change emotions; e.g., seeking advice/comfort from others), some surprising patterns emerged with social context variables, though with small effect sizes. For example, seeking advice/comfort from others was more likely to help when a user was around strangers and less likely to help when a user was around classmates. While it might be expected that this would be a more helpful strategy when a user was around friends, a romantic partner, or family, none of these contexts had significant effects on this strategy, suggesting that it would be interesting to see whether these patterns would persist if these recommendations were deployed to users. One interesting question that cannot be answered with the current study is how people sought social support; it is possible that people texted or called a friend when they were around strangers, so they may have still used friends to regulate even when those people were not immediately available in their physical environment. Strategies were generally predicted to be more effective when users were interacting with others than when they were alone or around others but not interacting with them, suggesting that the involvement of others might help users regulate effectively.

The effect sizes in Table 3.3 have some implications for the design of future studies. In order to minimize participant burden and to maximize the usefulness of the data collected, a researcher might focus more on collecting the most important features (i.e., those that have the largest effect sizes in predicting the 10 strategies considered in this chapter). The most important features were appropriateness of the time to be interrupted; energy, standard deviation, and mean of acceleration; whether participants had recently been in a car, walking, or stationary, their social

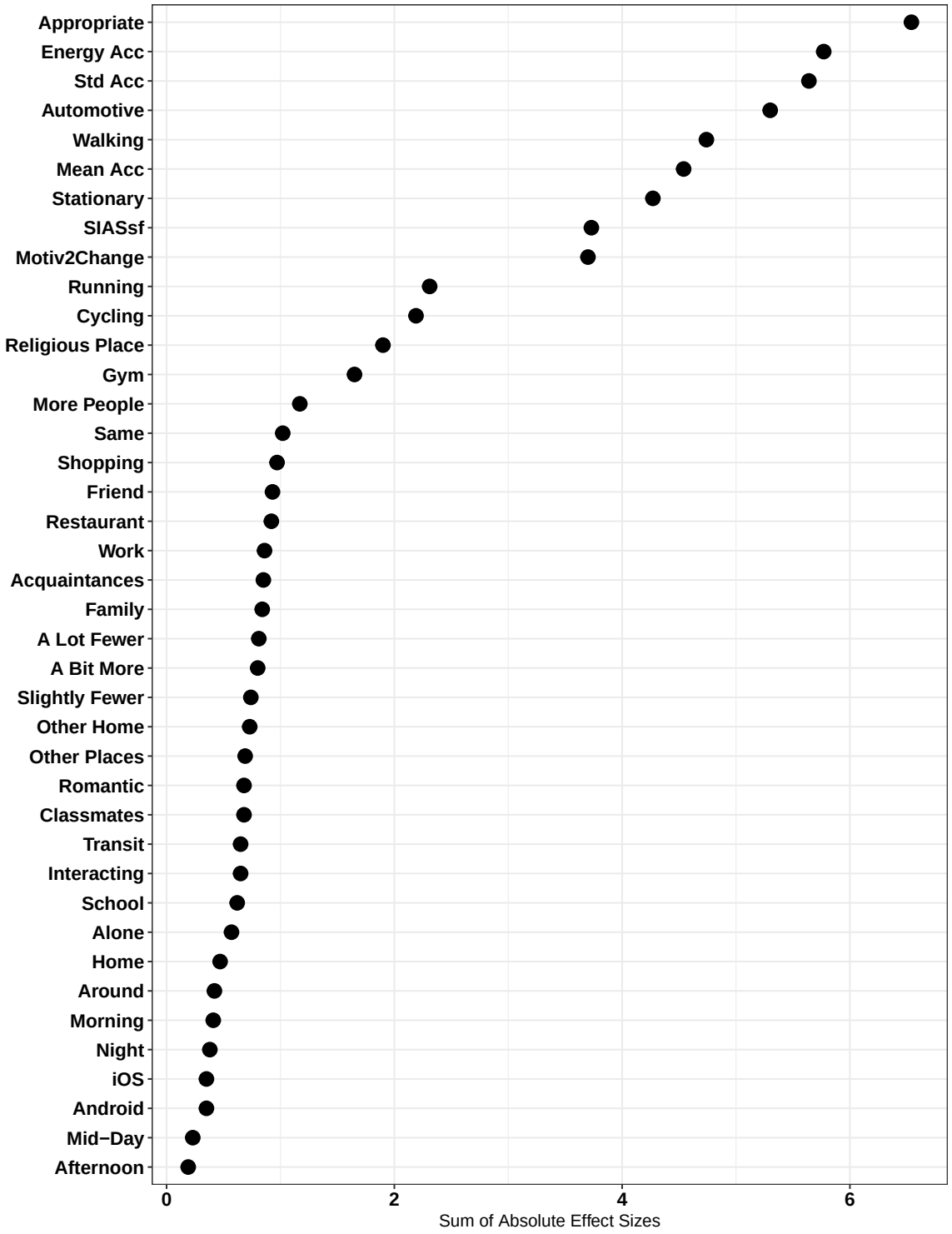


Figure 3.2: Ranking of contextual variables showing most critical features in determining the effectiveness of strategies. The ranking is based on the sum of absolute values of effect sizes.

anxiety symptom severity, and their motivation to change their thoughts/feelings. These important contextual features are all either continuous or discrete variables with many possible values; the less important contextual features are binary. This suggests that future researchers may aim to maximize the predictive value of their contextual variables by considering contextual variables with more variability in their values, as opposed to binary variables. Many of these more important contextual features also reflected movement, so future researchers may wish to preferentially include sensors that capture information about motion.

The current algorithms work to maximize *short-term* perceived effectiveness of regulating emotions, given the ER strategy attempt and effectiveness rating are reported close in time. However, psychologists have noted that both short-term and long-term regulation are important, with strategies differing in their effectiveness at different timescales [Freitas and Salovey, 2000]. For example, if you are anxious about an assignment due in a few days, watching TV might make you feel better for 30 minutes but leave you feeling anxious the next day, whereas tackling the issue and starting the assignment might feel worse for the next 30 minutes but make you feel better the next day. While CMAB optimizes for short-term ER effectiveness, evaluating the algorithms for longer-term effectiveness, examining a wider range of ER effectiveness indicators, and examining the algorithms in more diverse samples may all be beneficial directions for future work. Another limitation of this work is that when this policy is deployed, a user will initially need to request an intervention before the most contextually effective strategy is suggested; ultimately, the goal is to be able to passively determine future emotional states and send interventions without the user's initiation.

3.8 Summary

In this work, we present a novel application for contextual bandits to learn contextually effective strategies for ER. Our approach is distinct from most existing work in health recommender systems in that we learn an initial policy that might have a

positive impact on user engagement when finally deployed, as well as on sample efficiency in the online setting. Our results demonstrate that an experimenter can use available observational data to learn the usefulness of a new intervention policy; this may provide an efficient way to generate hypotheses that can later be tested in (resource intensive) randomized clinical trials. Given that ER is impaired across many mental illnesses, this work has the potential to enhance the availability of scalable interventions that can be used in daily life for many people.

4 | Subgroup-Based Emotion Regulation Policy Generation

Many people suffer from mental health disorders such as social anxiety and depression. Studies have shown that about 12% of Americans will experience social anxiety disorder at some point in their lifetime and, of these, 80% will not receive any treatment [Kessler et al., 2005a,Grant et al., 2005]. Clearly, there is a need to improve access to mental health treatment. In recent years, smart devices, such as mobile phones and watches have become commonplace and part of people's regular activities of daily living. This creates a new avenue for delivering personalized mental health interventions to users when and where they need them most.

Emotion regulation (ER), or the process by which we influence the emotions we express and experience [Gross, 1998], is essential for well-being but impaired in people with many forms of mental illness, including social anxiety disorder. Our aim in this chapter is to extend recent work using contextual bandits to create health recommender systems for ER (i.e., systems that optimize recommendations for ER strategy use), by investigating the hypothesis that there are subgroups of socially anxious individuals with different responses to ER strategies. In other words, we want to improve the utility of the generalized ER intervention policy for different subgroups in the cohort.

While it is ideal to run an online version for intervention recommendation for ER, akin to micro-randomized control trials [Klasnja et al., 2015], this is not always practical and requires extensive resources, so is often not a cost-efficient step

when developing new systems. A viable alternative, which we implement in this work, is to learn an offline or off-policy ER recommender system using retrospective data, which can then be adapted to warm-start an online version of the recommender policy. Previous work using contextual bandits to build a health recommender system for ER strategies have largely focused on learning a generalized context dependent policies [Ameko et al., 2020, Beltzer et al., 2020], also outlined in Chapter 3. In this chapter, we extend the contextual bandit ER recommender policy to subgroup level systems discovered using the k-means clustering algorithm.

4.1 Related Work

For many years, ER strategies have been categorized as adaptive and non-adaptive [Aldao, 2013, Bonanno and Burton, 2013]. But recent studies suggest that these strategies are neither just adaptive or non-adaptive, instead, the effectiveness of any strategy depends on the context in which it was applied [Doré et al., 2016]. Consequently, several findings have indicated that contexts such as demographic attributes like gender and culture might have a differential effect on the impact of emotion regulation strategies in a person's experience of emotions [Nolen-Hoeksema and Aldao, 2011b, Ford and Mauss, 2015]. A more recent paper that combined methods from machine learning and mobile sensing technology to build a context dependent ER recommender policy formally operationalized the idea that ER effective -ness is context dependent [Ameko et al., 2020, Beltzer et al., 2020]. This work builds on these studies to improve the utility score by learning specialized context-aware ER policies for subgroups in the population.

Health recommender systems create a viable way to scale up access to health care by leveraging smart devices with embedded sensors to help monitor and deliver timely intervention to users. A notable example of health recommender systems is, myBehavior, a mobile app that tracks user's physical and dietary habits, recommends personali -zed suggestions for a healthier lifestyle [Rabbi et al., 2015] using multi-armed bandits. Cheung et al. [Cheung et al., 2018] created a mobile app called

IntelliCare, which consists of a suite of 12 individual apps recommended as 'treatments' for managing depression and anxiety symptoms. Yang et al. [Yang et al., 2018] created a mobile health recommender system that integrates depression prediction and personalized therapy solutions to patients with emotional distress. In their system, personalization is realized using external factors related to depression, including family life, external competition, interpersonal relationship, self-promotion burden, economic burden, work pressure, individual personality, coping style, and social support, which are assessed using mobile questionnaires. These mobile health efforts are consistent with a mobile intervention framework called Just-in-time adaptive interven -tion (JITAI) [Nahum-Shani et al., 2017].

The JITAI framework is characterized by major branches of research, specifically, the timing of intervention delivery and choosing the best intervention strategy to deliver. A considerable amount of research focus has been directed towards optimizing for the best timing to deliver an intervention (e.g., predicting stressful moments linked to emotional eating [Rahman et al., 2016]). By contrast, our work focuses on identifying the most effective ER strategies based on a person's context and subgroup in the population. While reinforcement learning is typically used to formalize the object of mapping context to the right ER strategy [Ameko et al., 2018, Beltzer et al., 2020], we use unsupervised learning methods to determine user subgroups at baseline.

Subgroup analysis as applied to health recommender systems has gained recent popularity in the literature, since there is typically very limited data to train a separate model for each person, and yet still maximize the relevance of the recommenda -tions to users at the early stages of deployment. This issue is particularly important in mobile health because of the risk of disengagement or attrition caused by making suboptimal interventions. Hassouni et. al. demonstrated in a simulation study that a cluster-based reinforcement learning for mobile health interventions procured significant advantages in term of policy effectiveness and data efficiency in the early stages of the experiment [el Hassouni et al., 2018]. Similar findings were made by Zhu and Liao, showing significant gains of subgroup based reinforcement learning over generalized approaches, and particularly showing that the approach is feasible in the off-policy setting for warm-starting future recommendations used in online

studies [Zhu and Liao, 2017]. In addition, Tomkins et.al., demonstrated advantages of intelligently pooling users together into similar groups within the online intervention design setting [Tomkins et al., 2021]. Similar to these works, we use a subgroup analytic method to train personalized recommender algorithms, and unlike previous work, also apply it in an offline setting on a real-world dataset.

4.2 Method

4.2.1 Study Design

The study consisted of N = 114 participants aged 18 years and older (M = 20.39, SD = 2.94, range: 18-34) from a US college and community. Enrollment in the study was determined by the Social Interaction Anxiety Scale (SIAS) [Mattick and Clarke, 1998] (scale range is 0-80), with a cuttoff point set at 29 and above to enroll participants experiencing moderate to severe social anxiety symptoms. The analysis excluded 4 participants due to missing data; specifically 1 participant did not report any ecological momentary assessment (EMA) survey data and 3 other participants did not report the effectiveness score of ER strategies which we use to train the recommender model. The remaining 110 participants consisted of: 81 female, 29 male (with no participant reporting a non-binary gender identity), 86 undergraduates, 11 graduates or professional students and 13 others. The participants were aged 18-34 with mean = 20.4 and SD = 2.98, and 82 participants reported their race as White/Caucasian, 21 Asian, 7 African American, 3 Middle Eastern, 3 Native Hawaian/Pacific Islander (these numbers exceed 110 as many participants self-identified as multiple races). The participants SIAS scores ranged from 29-73 (mean = 46.68 and SD = 10.39). Although the full SIAS score was used for the enrollment process (as a way to be consistent with the literature), the sum of the straightforwardly-worded items was used for the analyses, as the straightforwardly-worded items have been shown to improve the psychometric properties of the SIAS score [Rodebaugh et al., 2007]. Participants were compensated for the five weeks of EMA based on how many EMA surveys they completed, ranging from $10 to $80.

During a baseline laboratory session, participants who consented to enroll in this study completed several questionnaires (some of which were analyzed in this study) and behavioral tasks. Subsequently, MetricWire, a smartphone app, was installed on all participants' personal smartphones. For the next five weeks, MetricWire was programmed to deliver six identical, randomly timed surveys through -out each day, with randomly timed survey prompts sent every two-hour block between 9am and 9pm (i.e., once between 9-11am, once between 11am-1pm, etc.). Each survey took approximately two minutes to complete. Participants were instructed to complete the survey as soon as possible upon receiving the notification. Participants received a reminder notification 30 minutes after the initial notification if they had not yet completed the survey, and the survey disappeared 15 minutes later. Participants were instructed to answer the survey with reference to the time of the initial survey notification. This instruction might introduce a small degree of recall bias into survey responses, but should enhance ecological validity by sampling a wide variety of situations in daily life, including situations in which it would be difficult to respond to a survey immediately (e.g., when a participant is taking an exam or in the middle of a meeting). Sensor data were also passively collected from participants' smartphones to capture their activity levels and GPS locations.

4.2.2 Clustering Dimensions

To assess the differences in emotion regulation effectiveness of population subgroups, we consider two categories of features, denoted as baseline and sensor-extracted measures, used to discover these subgroups where users are assumed to be similar. Each set of features within a category produces a multi-dimensional profile vector to characterize each participant in the study cohort. We rely on previous research findings to select each feature which we describe below.

Baseline Measures

Demographic Features

Emotion regulation strategies have been found to vary across people of different genders and ages [Nolen-Hoeksema and Aldao, 2011b]. Furthermore, culture affects both emotion regulation motivation and well-being [Ford and Mauss, 2015]. Given these findings and the ease of assessing demographic features, we consider including demographic features assessed at baseline that have sufficient variability within our sample, which may include sex, age, ethnicity, race, native language, and household income level.

Symptom Severity

Although emotion dysregulation is a transdiagnostic process in psychopathology, use of specific emotion regulation strategies varies across psychological disorder [Aldao et al., 2010b]. For example, a person with a substance use disorder might be more likely to turn to drugs or alcohol to regulate their emotions, while a person with major depression might take a nap to deal with a similar emotion. As such, we create clusters based on baseline symptom severity for several psychological disorders (namely, social anxiety disorder, generalized anxiety disorder, major depressive disorder, and alcohol use disorders). The details about the features extracted from these measures are included in Table 4.1.

Emotion Regulation Processes

Self-reports EMA surveys of specific styles of emotion regulation and dysregulation (e.g., difficulties engaging in goal-directed behavior, impulse control difficulties) predict in-the-moment emotion regulation strategy choices and the effectiveness of those strategies [Daros et al., 2020]. Based on these findings, we include features on self-reported trait emotion regulation styles and difficulties in creating participant profile using the subscales of two emotion regulation questionnaires. These emotions regulations questionnaires include, Difficulties in Emotion Regulations Scale (DERS)

[Kaufman et al., 2016] and Emotion Regulation Questionnaire (ERQ) [Gross and John, 2003]. Details of the subscales are outlined in Table 4.1.

Sensor-derived Features

Passive Features

Passively sensed data from smart devices can be collected unobtrusively from a user to capture momentary traces of a person's behavior at a granular level. Data such as accelerometer and GPS provide activity and location data shown to be highly associated with important mental health issues, including social anxiety symptoms and depression [Boukhechba et al., 2018,Ameko et al., 2018]. In this work, we propose to use these passive data features including activity types (e.g., walking, automotive etc..) and semantic location (e.g., Home, School etc..) to capture a user's daily routine over the first 7 days of the study. These features specifically measure the distribution of users' physical activity level; for example, a user may spend 60%, 30%, and 20% of his/her time at home, in school and at other places, respectively.

4.2.3 Counterfactual Estimation Methods

Similar to Chapter 3, we formulate learning warm start ER policy using offline contextual bandits in which our policy is derived from measuring contextual interven -tion effects. Contextual bandits is a variant of reinforcement learning which leverages additional information about the world (e.g., context) for decision making. Formally, given an agent that interacts with an environment over a finite number of steps, denoted T, the agent perceives contextual input signal, x, and chooses an action $a \in a_{i}{}_{i=1}^{k}$ that maximizes its reward. In this application, the perceived environment consists of a combination of features from passive and EMA survey data capturing information about the environment of the user. Examples include semantic locations and activities.

In the offline policy training setting, the featurized observational data (e.g., passive and EMA) collected from the smart device is assumed to be generated from

Construct	Scale	Scoring
Symptom Severity		
Social Anxiety Symptoms	Social Interaction Anxiety Scale (SIAS), a 20 item questionaire [Mattick and Clarke, 1998]	Sum of items except question 5,9 and 11. **(SIASsf)**
Depressive Symptoms	Patient Health Questionnaire (PHQ-8), an 8 item questionaire [Kroenke et al., 2009]	Sum of items **(PHQ)**.
Generalized Anxiety Symptoms	Penn State Worry Questionnaire - Ultra Brief Version [Berle et al., 2011].	Sum the items **(PSWQ)**.
Emotion Regulation Processes		
Nonacceptance of emotional responses	Difficulties in Emotion Regulation Scale – 18 Item Short Form (DERS-18) [Kaufman et al., 2016]	Sum DERS7, DERS12, DERS16 **NonAcceptance**.
Difficulty engaging in goal-directed behavior		Sum DERS8, DERS11, DERS13 **(Difficulty)**.
Impulse control difficulties		Sum DERS9, DERS 14, DERS 17 **(ImpulseDiff)**.
Lack of emotional awareness		Sum DERS1R*, DERS4R*, DERS6R* (reversed scale) **(NotEmoAwa)**
Limited access to emotion regulation strategies		Sum DERS10, DERS15, DERS18 **(LimitedERAccess)**.
Lack of emotional clarity		Sum DERS2, DERS3, DERS5 **(NoEmoClair)**.
Reappraisal	Emotion Regulation Questionnaire (ERQ) [Gross and John, 2003]	Sum ERQ1, ERQ3, ERQ5, ERQ7, ERQ8, ERQ10 **(Reappraisal)**.
Suppression	ERQ	Sum ERQ2, ERQ4, ERQ6, ERQ9 **(Suppression)**

Table 4.1: A description of the self-reported baseline features extracted for clustering. **(Bold)** names are feature names used in the rest of the chapter.

an unknown policy called the behavior policy, and denoted π_b. Learning a target policy, denoted π_e from this data consists of solving the objective function,

$$\pi_e^* = \underset{\pi_e \in \Pi}{\operatorname{argmax}} \mathbf{V}^{\Pi}. \tag{4.1}$$

where $\mathbf{V}$ denotes the value of a policy and Π, the policy class. The policy value given π_e^* is

$$\mathbf{V}^{\pi_e^*} = \mathbf{E}_{(x,r)\sim D}[r_{\pi_e^*}(x)]. \tag{4.2}$$

We consider a linear policy class which are efficient for training and easy to interpret. We apply importance sampling techniques that use a form of weighting scheme denoted as $\frac{\pi_e(a_i|x_i)}{\hat{\pi}_b(a_i|x_i)}$ in context x_i to correct for the distributional shift between the target and behavior policy in order to have an unbiased estimate of the target policy value [Dudík et al., 2014].

There are three main value estimators that lie at the core of offline policy training and evaluation within the contextual bandit framework; namely, the Direct method, Inverse Propensity Weighting, and Doubly-Robust. In this work, we use the Offset Tree, a prominent algorithm using doubly robust estimation which is shown to work consistently on observational data [Ameko et al., 2020, Beygelzimer and Langford, 2009].

Propensity Score Estimation

As noted above, the behavior policy that generated the data is unknown and needs to be estimated from the observed data. This is achieved by estimating propensity scores which represent the likelihood of choosing different strategies in different contexts. Propensity scores also serve to reduce multivariate contextual data [Rosenbaum and Rubin, 1983] into one-dimensional scores such that treatment group distributions are matched. The goal of the propensity scores is to create a pseudo-population where contextual distributional overlap across strategy groups is sufficiently achieved.

Ensuring overlap in the strategies with respect to the propensity scores reduces

the possibility of extreme values in the estimation step, given that the estimation approaches depend on the score denoted $\hat{\pi}_b(a|x)$. Estimation methods such as logistic regression have typically been used but they are limited in expressiveness due to their linearity assumption [McCaffrey et al., 2013]. By contrast, there are non-parametric machine learning models developed to add more flexibility in order to model more complex data which is expected from human generated data. An example of a non-parametric model is Generalized Boosted Models (GBM). GBM estimation uses an iterative process with multiple regression trees to capture nonlinear relationships between strategies and context variables without over-fitting the data. We implemented GBM propensity score estimation in our analysis using the R package *twang* [Ridgeway et al.,]. We used the absolute standardized mean difference [Stuart et al., 2013] as the stopping criteria over 5000 iterations.

4.3 Data Processing Methods

To model the data using our proposed methods, several decisions were made with regards to data pre-processing and setup for training the recommender policy. The decisions made include how features were chosen and processed for analysis, how missing data was handled and what imputation methods were used. Finally, we describe the model framework and how the data maps into the different components of the contextual bandit learning algorithms.

4.3.1 Feature Extraction

Accelerometer Features

We extract features pertaining to infer user movement patterns from the raw accelero -meter derived from the smartphone. The raw data comes in the x, y, z coordinate form of acceleration over time. To remove the effect of sensor orientation, we derived an orientation invariant measure z user movement using $z = \frac{1}{3}\sqrt{x^2 + y^2 + z^2}$. We derived standard accelerometer features such as mean, standard deviation and average

energy [Hachiya et al., 2012]. Accelerometer features have been as predictive measures of mood and affect [Jaques et al., 2017, Ameko et al., 2018] suggesting that this could explain changes in our participants reported measure of affect and perceived effectiveness of strategies. In total, we extracted 3 features from the accelerometer data.

Semantic Location Features

In addition to movement patterns, we also extracted semantic locations directly from participants through the EMA surveys. At each survey prompt, the user is asked to provide a location label with reference to when the prompt went off. Users are provided a list of options of semantic locations which were seen to be the most prevalent from previous studies. The locations include the Gym, Home, being in transit from one location to another, Other's Home, Religious Place, Restaurant, School and Shopping locations. When participants were in places other than the options provided, they chose other locations and had the option to input a text describing their location. In our analysis we grouped these other locations into one feature as the collected details of the locations were extremely noisy and varied. Location has similarly been found to be associated to a person's mood [Jaques et al., 2017].

Activity Types Features

We also describe activity modes such as cycling, running, automotive, stationary and walking. These activities are directly determined using the labels provided directly by MetricWire. The features are treated as categorical in our analysis. Physical activities have been recognized as influential in a person's experience of negative and positive affect. Researchers have empirically shown that active physical activity leads to an improved level of affect compared to sedentary lifestyle [Guszkowska, 2004]. The differential effect that might ensue from these activities is encoded in our analysis by including them as context variables for the recommender system algorithm.

Time of Day

Time including time of day and day of week are known to have differential effect on a person's mood. Studies have shown that people generally experience higher positive affect in the afternoon or sometime around midday with respect to time of day [Egloff et al., 1995]. We include features of time of day (e.g., morning (9-12PM), midday (12-3PM), late-afternoon (3-6PM), and night (6-9PM)) in our analysis.

Social Partners in Context

Studies have shown a consistently robust link between daily positive mood and social events (e.g., parties, leisure time with friends and family, social eating and drinking events) [Watson et al., 1992, Clark and Watson, 1988]. Given how important this variable appears to influence a person's affect in studies designed to obtain feedback over weekly periods, we hypothesize that the effect might be more significant at the hourly level setting.

Social Preference of Users

Participants' social preference may provide clues as to the effectiveness of social ER strategies; strategies like seeking social support or doing something fun with others may lead to greater improvements in mood when participants wish they were around more people.

Social Interactions in Context

Studies have shown that people with high depressive symptoms tend to react more strongly to positive and negative social interactions [Steger and Kashdan, 2009]. We capture user interaction contexts using data from our EMA surveys in which users were asked to provide information about their social interaction states, whether alone, with others but not interacting or with others and interacting.

4.3.2 Approaches to Handle Missing Data

Data collection in mobile health applications are usually fraught with missingness due to several factors such as power issues, software bugs, user non-response, user interruptions (i.e., manually switching devices), or data that is missing due to asynchro -nous sampling rates from different sensors. Missing data is categorized as missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR). MCAR refers to a missingness situation where the missing data is assumed to be unrelated to any variable both observable or unobservable, in other words, it happened by chance. However, this assumption is rarely valid in the real world since most systems are linked together and depend on each other. On the other hand, MAR is the situation where missingness is assumed to be related to an observed factor which is not necessary related to a measurement under study, in our case, ER effectiveness. As a result, imputation for missing data can be done by adjusting for observed data. The MNAR occurs where there are assumed to be counfounders that affect missingness as well as the measurement under study. For example, a user might turn off their device as a result of feeling extremely low in desire or motivation to do something about their present affect. Here, the person's internal motivation can be a confounder that might be difficult to measure in most studies. MNAR situation is often complicated by the inability to statistically test for the existence of such confounding variables. In this chapter, we adopt the MAR assumption and impute missing data for accelerometer and activity types, with missingness at 65% and 68% respectively, using multiple chained imputation. The multiple chained imputation algorithm works by iteratively imputing missing values in the input data by regressing each column of missing data on the rest of the imputed column from the previous step until convergence or a pre-specified number of iterations is attained. This method, has been shown to work well in mobile health applications [Rashid et al., 2020]. In this chapter, we used the MICE statistical package in R with 5 iterations.

4.3.3 Discovering Clusters

To discover the subgroups in the study sample, we considered two categories of features; using passive, sensor-extracted features describing user routine, as well as, baseline self-reported measures collected on users at the laboratory session. Each of these categories can be useful in different settings depending on what data is available; for instance, if there are baseline measures then the policies based on these measures can be easily adopted, whereas, in the setting with no baseline measures, one can adopt the policies trained based on passively-extracted features for clustering. Each clustering category formed a multi-dimensional feature vector depicting a profile for every user in the cohort. The passive features used to create user profiles were extracted using the first 7 days of users' data collection period. As a results, ten users with less that 7 days of data were excluded from the clusters formed within this category. The passive features are comprised of contextual-information extracted and semantic geographical locations that users confirmed via the random EMA surveys from the mobile app. From these features we extracted the user's daily routine by using the empirical frequency over the features. For example, for location features, an arbitrary user might spend an average of 50%, 30% and 20% of their time at Home, School and Gym for the first 7 days. The baseline self-reported measures, as summarized in Table 4.1, were summed over sub-scales for each questions related to the baseline measures considered for clustering. We used the k-means algorithm with the Euclidean distance metric to find similarity among users. We set our K to optimize for the silhouette score [Rousseeuw, 1987].

4.3.4 Model Framework

Our contextual variables capture the user's state around the time they used a strategy as reported in the EMA. A combination of the contextual variables allows us to provide contextual recommendation for ER strategies. For example, given that a user is at home in the evening with a trait social anxiety level of 30, we might recommend tackling issues head on if our algorithm predicts it to be the most effective strategy.

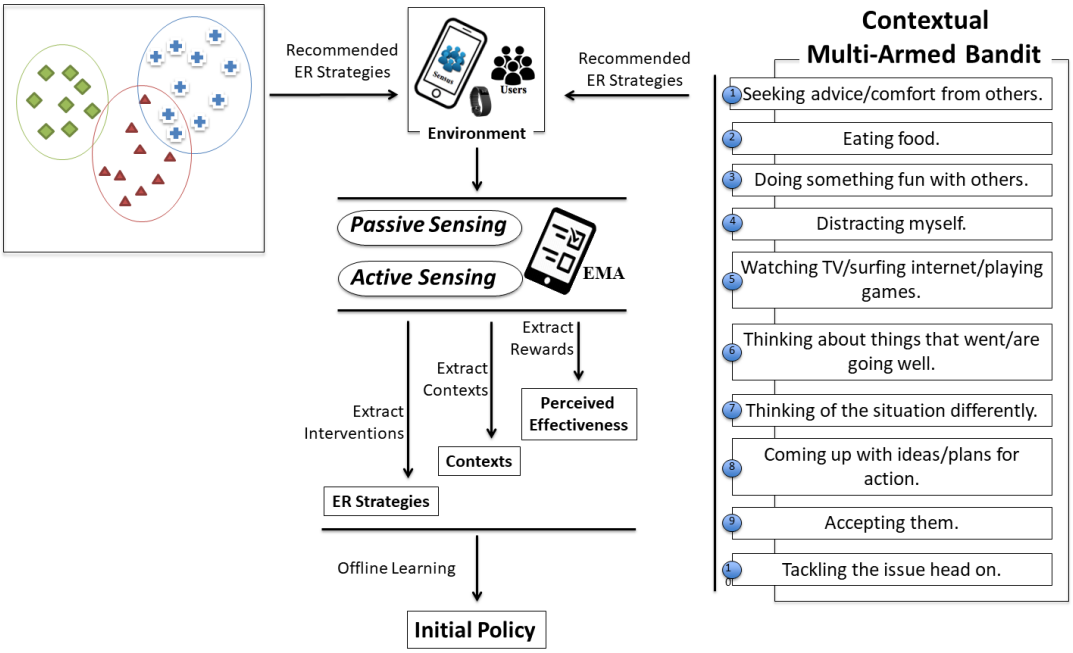


Figure 4.1: Model Conceptualization for Subgroup-based Emotion Regulation

The actions in our formulation are the top 10 most frequently used adaptive strategies, which are shown in the contextual multi-armed bandits in Figure 4.1. It is notable that there are multiple ways to reduce dimensionality of the ER feature space, and we ultimately explored alternative approaches in other analyses. However, we chose to focus on this subset of strategies as they are mostly considered healthy strategies (i.e., they tend to be associated with positive health consequences, unlike a strategy such as using alcohol or drugs to change one's feelings) and these were also the most frequently reported strategies in our learning data.

The reward signal needs to reflect the effectiveness of the chosen strategy in the given context at helping to manage the participant's emotion. In our data, participants reported the perceived effectiveness of their ER attempt on a scale of 0-10. We dichotomized this outcome measure, both for problem simplification and leveraging existing theory [Strehl et al., 2010], to define a reward signal for the agent. Our threshold was defined as the average of effectiveness scores across all users within each discovered subgroup, or the subgroup grand mean. Specifically, given a subgroup indexed c, let $O_k(x_i, a_i)$ denote the immediate effectiveness of the chosen ER strategy at time i in context x_i, we have the grand mean as

$$\hat{O}_c = \frac{1}{N_c} \sum_{i=1}^{T} O_k(x_i, a_i), \tag{4.3}$$

. Where N_c denotes the size of sample within the indexed subgroup. The reward signal for each context x and action a is thus defined as:

$$r(x, a) = \mathbf{1}_{\{O_c(x,a) > \hat{O}_c\}}, \tag{4.4}$$

where $\mathbf{1}$ is an indicator function that returns 1 when the condition is satisfied, and 0 otherwise.

4.3.5 Training Models and Evaluation Methods

Training Model

We used logistic regression as the value function approximator to map contexts to rewards for each ER strategy within the Offset Tree algorithm in our policy training step. The offset tree is a doubly robust estimation method that learns several binary regression models with propensity weighted reward offset by a constant factor to control variance of estimation. More details can be found in Beygelzimer et al. [Beygelzimer and Langford, 2009]. We compare the performance of the trained policy per subgroup against the global policy which is generalized for the entire cohort.

Model Evaluation

Given the selection bias in the test data, we evaluate the performance of the different recommender algorithms using the importance sampling based approached named the trimmed inverse propensity weighting (tIPW), also called the mean reward. This is specifically defined for a test set of size T as,

$$V_{tIPW} = \frac{1}{T} \sum_{i=1}^{T} \frac{\pi_e(a_i|x_i)}{max\{\hat{\pi}_b(a_i|x_i), \tau\}} r_i. \tag{4.5}$$

The parameter τ is a nuisance parameter and that can take values $\tau < 1/k$, where k is the number of strategies, according to lemma 3.1 of Strehl et. al., [Strehl et al., 2010].

4.4 Results

4.4.1 Experiment Details

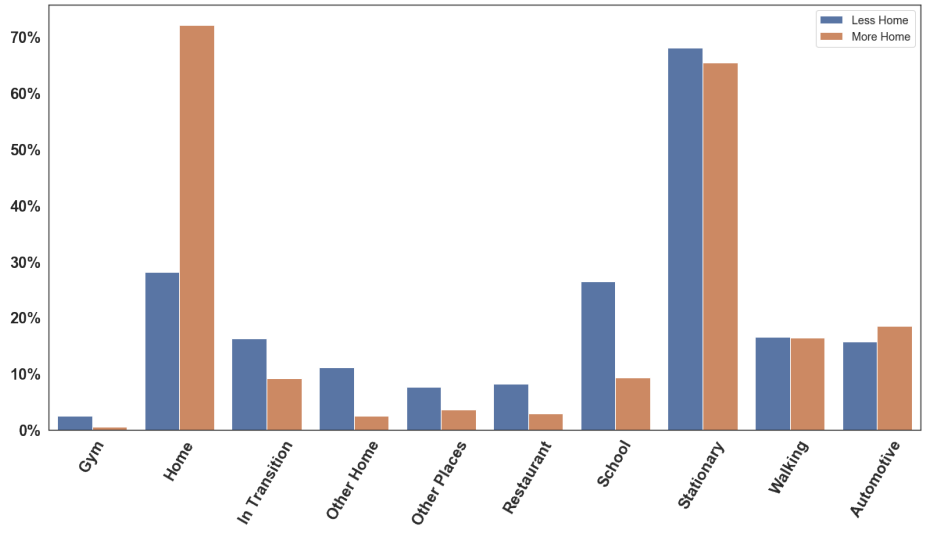
Our experimental setup is similar to the procedure outlined in Chapter 3 [Ameko et al., 2020] in terms of data processing and the learning of the propensity scores.

We modified the approach by learning a separate propensity scores for each subgroup discovered to reflect our belief that users in the same cluster are expected to have similar behavior, and we maintained the reward definition to allow consistency across clusters.

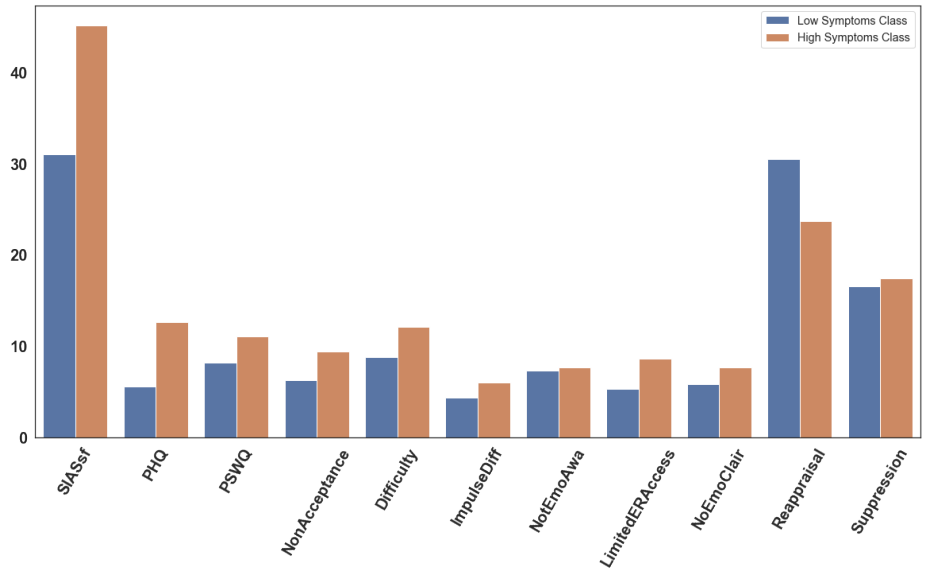
In addition to the passive features and baseline self-reported features used to cluster participants, we also considered demographic features such as age, income level and race in our experiments, but the cohort is homogeneous along these features and lack enough variability for clustering. We conducted ablation studies by dropping demographic features from the clustering algorithms and realized no change in the silhouette score, therefore these features were excluded from our final analysis. So although we considered using a k-prototype [Huang, 1998] clustering algorithm to handle the mixture of continuous and discrete variables, we ended up using the k-means since all our clustering features were continuous. Figure 4.3 shows the cluster centroids for each cluster by category. Each cluster depicts the prototypical profile of a user belong to the subgroup. Four subgroups emerged in our analysis. In Figure A.3b, there were two subgroups that we entitled *Less Home* and *More Home* due to the clear distinction in the percentage of time spent at Home in each group. Similarly, in Table A.3a, we have two clusters, *Low Symptoms* and *High Symptoms*, with the clear distinction being the symptom severity.

We use the Offset Tree algorithm to train a cluster specific ER recommender policy and evaluate the trained policy using the trimmed inverse propensity weighting 4.5 as done in Chapter 3 [Ameko et al., 2020]. We vary the trimming parameter τ between $[0, 0.1)$, specifically we use $0.0, 0.02, 0.05, 0.07$ in our experiments, as shown in Figures A.3a and A.3b to control variance and reduce uncertainty.

Our experiment results comparing the global model to the subgroup models from the two categories of clustering features show an improvement over the global baseline for three out of four clusters discovered. Figures A.3a and A.3b show that mean reward from the clusters compared to the global at each trimming parameter τ. We performed a Tukey HSD (Honestly Significant Difference) test to understand the statistical significance of these models at $\tau = 0.02$ (e.i., where model difference

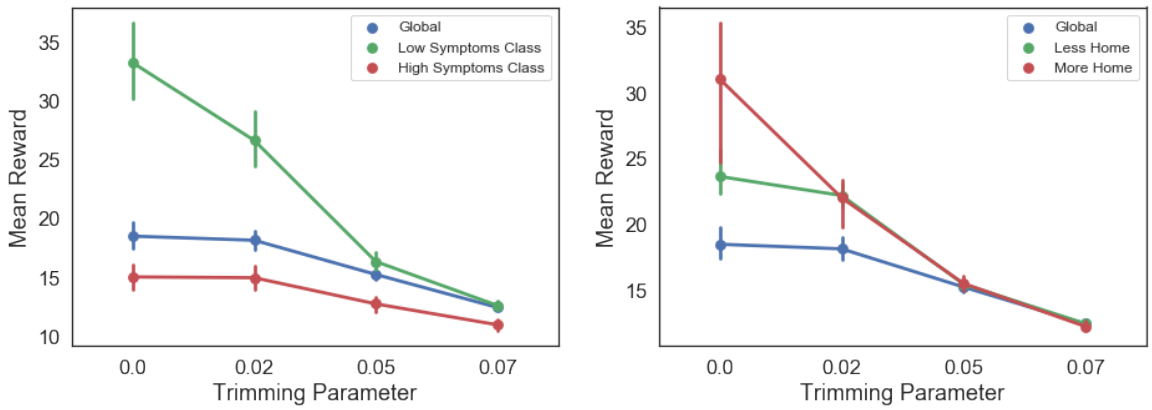


(a)



(b)

Figure 4.2: Compared mean policy rewards of subgroups based on passively sensed features of subject routines versus global policy A.3b, and baseline self-reported measures subgroups versus global policy A.3a



(a) **(b)**

Figure 4.3: Compared mean policy rewards of subgroups based on passively sensed features of subject routines versus global policy A.3b, and baseline self-reported measures subgroups versus global policy A.3a

Groups	Subject Size	Sample Size	Policy Mean Reward
Global	110	6259	18.11 ± 2.34
Less Home	56	2498	**22.17 ± 2.22**
More Home	44	2106	**21.99 ± 5.16**
Low Symptoms	50	2719	**26.58 ± 5.97**
High Symptoms	60	3540	14.93 ± 2.64

Table 4.2: Summary statistics on various clusters. The figures in bold indicate a significant uplift over the global baseline at $\alpha = 0.05$.

are clearer with the least standard deviation). The details of each cluster in terms of sample and subject size with the mean reward with 95% confidence interval at $\tau = 0.02$ are summarized in Table 4.2.

4.5 Discussion

Our subgroup analysis led to improvements over the global policy for both routine-based clusters and the lower symptom severity self-reported baseline subgroup. This suggests that clustering generally tended to improve the recommendations of the algorithm, especially for participants with less severe symptoms or when passive

features were used for clustering. However, for the higher symptom severity subgroup, the cluster policy did not perform significantly worse than the global policy. These results suggest that in creating a warm-start policy for a recommender algorithm, both self-reported baseline measures and passive features are useful. However, passive features seemed to be similarly useful for tailoring warm-start policies for both clusters, whereas self-reported baseline features were particularly useful for those with less severe symptoms, and not as useful for those with more severe symptoms. Features like the locations people visit may offer important clues into their common contexts and behavioral patterns that may translate into different ER recommendations.
Similarly, at lower levels of symptom severity, tailoring a warm-start policy may offer more benefits than at higher symptom severity. This might be because participants with more severe symptoms have more complicated profiles of ER processes, difficulty enacting ER strategies, and more pervasive impairments, such that even with tailoring, it is difficult to pinpoint one policy that tends to help this group.

The most notable difference between the recommendations for the routine based clusters is that the cluster who visited varied locations (e.i., *Less Home*) was more frequently recommended distraction, whereas the cluster who stayed home more (e.i., *More Home*) was more frequently recommended acceptance, see Figure 4.4. These two ER strategies take very different approaches toward anxious thoughts: distraction involves shifting attention away from thoughts, and acceptance involves continuing to pay attention to thoughts, but with a less judgmental stance. It is possible that these recommendations capitalize on the strengths or typical regulation styles of each cluster (i.e., the varied location cluster maybe good at engaging in activities that distract them, and the home cluster may be good at paying attention to their thoughts).

On the other hand, the most notable differences between the clusters defined by the baseline self-reported features is that the low symptom severity cluster (e.i., *Low Symptoms*) is more frequently recommended to do something fun with others, whereas the high symptom severity (e.i., *High Symptoms*) cluster is recommended to accept their thoughts/feelings and think about positive things, see Figure 4.4. We will not speculate as to what the high symptom severity cluster's policy may

reflect, as this policy did not outperform the global policy. For the low symptom severity cluster, however, doing something fun with others show up as a more highly recommended strategy in the cluster vs. global policy again because of capitalizing on strengths: due to their less severe symptoms, this cluster may have stronger social contacts whom they might invite to do something fun, whereas this might be less of an option for participants with greater social impairment and higher symptoms.

One limitation of this study is with the sample, both in terms of coverage and size. As noted earlier, the data is collected from a college student group and it is skewed towards a few population subgroups, hence limiting it's replicability in a larger and more representative population. Besides, the current subgroup analysis that requires a hard assignment of participants to a cluster led to significant improvements on the global policy, exploring a Bayesian hierarchical model to allow users to share membership in several clusters might lead to further improvements in policy perform -ance. Also, in this chapter, we optimize for immediate effects of strategies, which in a way mitigates the effects of noise in the reward; but emotion regulation is believed to have both short-term and long-term effects [Freitas and Salovey, 2000]. Consequently, modeling the long-term effect of ER strategies in a Bayesian hierarchical framework is an interesting direction to explore in order to overcome the current limitations of the trained ER policy in this work.

4.6 Summary

In this chapter, we present a subgroup analytic method for building a warm-start policy for emotion regulation strategies. We leverage historical data collected from baseline and smartphones and use offline contextual bandits to train and evaluate the policies. Our approach builds on previous work by demonstrating improvement in policy performance, therefore having the potential to further increase user engagement in future confirmatory studies. Given the importance of emotion regulation to human mental health, this work is one more step made towards the vision of increasing mental health treatment access to patients by leverage smart and ubiquitous devices.

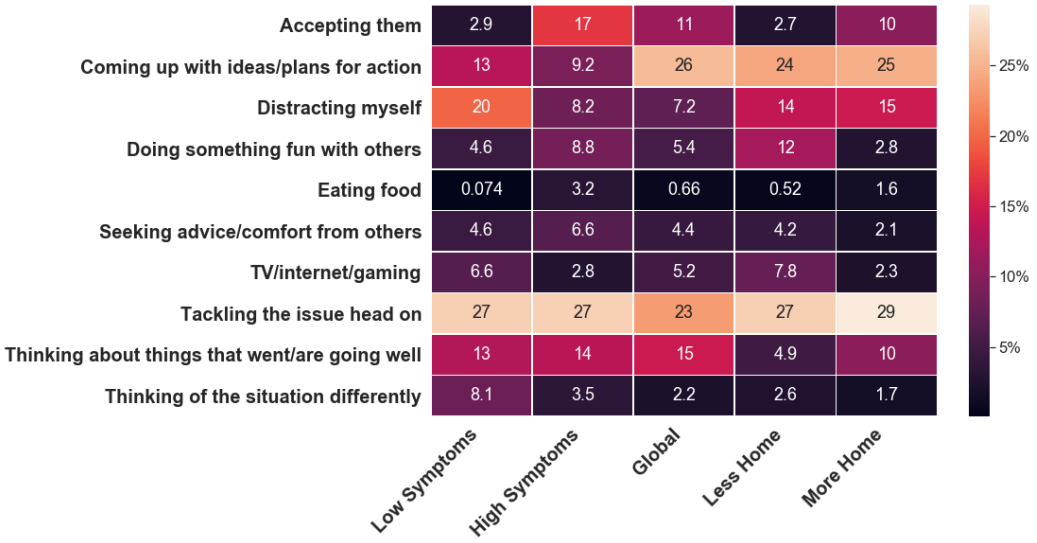


Figure 4.4: Policy distribution by cluster/subgroup marginalized over contextual variables to highlight the difference in recommendations made by each policy. We also compared the learned policies against the observed for more context.

5 | Conclusions & Future Work

Scalable access to healthcare has become more feasible with the increase ubiquitous devices such as smartwatches and smartphones. We can leverage mobile sensing streams and computational models to deliver more personalized and timely interventions. This work contributes methods for learning personalized mobile health outcomes and interventions in-the-wild. Using data from multiple modalities, including, mobile EMA and multimodal sensors, this dissertation addresses these research problems and makes the following contributions:

- **Personalized Prediction of Affect:** We used passively sensed data to characte -rize each participant profile in the study. These profiles were used to cluster participant into homogeneous groups based upon which we learn a prediction model for their self-reported affect in the wild. We used a collection of models including Gaussian processes, Lasso linear model, Random Forest and Support Vector Machines to demonstrate that group based predictions of affect performed better that a group independent model. We also, show that while learning models for each individual participant might be promising, there needs to be more data collected for this approach to be a viable option based on the statistical significance.

- **Generalized Warm-Start Intervention Policy from Mobile Digital Data:** We have developed a novel method to learning and evaluating a warm start intervention policy from historical digital data. This addresses the problem of cost involved in designing new intervention studies and mitigate the risk

of attrition by learning to provide useful interventions to users early on in a study. We leverage techniques from contextual bandits and causal inference to formalize the problem of learning warm-start policy which unlike existing work, uses real world observational data for a real mental health problem that has a wide impact on several psychopathologies. Essentially, this work derives an intervention policy for emotion regulation for any user from a moderate to high social anxiety disorder. This approach is personalized by the context of the user and lays the foundational step towards a subgroup relevant policy.

- **Subgroup Personalized Warm-Start Intervention Policy from Digital Data:** This final contribution addresses the limitation of the above intervention policy learning work by construction a subgroup relevant policy where similar patients in terms of baseline psychological measures or daily habits as captured by passive data are grouped together and provided a specialized warm-start policy. We characterized users from their baseline self-reported psychological measures and daily routines from passive measures and used k-means to cluster them into homogeneous groups. We show that by learning an intervention policy for each group, we significantly improve upon the group independent policy from our previous work.

5.1 Future Work

There are several interesting future directions where this body of work can be extended. A key limitation that cuts across all our studies especially for affect prediction is the relatively small amount of data across few participants. There is an opportunity to further validate, and evaluate the generalizability of our approaches by collecting a larger sample with more diverse participants.

Our algorithms in Chapters 3 and 4 use contextual bandits to model the effect of emotion regulation strategies both because of the limit in the data available and the mitigation of the effect of noisy observations in the reward as outlined in

subsection 1.4.1. Future work can consider leveraging a larger data set to learn a full-reinforcement learning policy where the long term effect of emotion regulations are accounted for in natural settings. Alternatively, there is an avenue to direct efforts towards user modeling in high fidelity simulations as is currently being pursued in areas of traditional recommender systems [McInerney et al., 2021, Rohde et al., 2018]. This will provide an benchmark for evaluating new intervention policies in mobile health before deployment.

Also, there is an opportunity to join together the techniques developed in Chapters 2 and 4 to allow for an automatic prediction of user affect and the deployment of timely interventions. This could take the form of a two-pronged model architecture with a nuanced variation like a feed-forward neural network where one head of the model predicts the affect score of the user and the other head determines the most useful intervention to deliver in opportune moments.

Finally, there is a exciting new direction to apply inverse reinforcement learning by training a reward function from demonstrations collected from less socially anxious individuals and using the trained reward function to guide the intervention policy for high socially anxious individuals [Abbeel and Ng, 2004]. This approach will help overcome the simplistic reward function defined in this thesis and provide much more benefits for long-term reinforcement learning approaches. Alternatively, inverse reinforcement learning can be a viable approach to learning individual behavioral routines from observed sensor data, as shown in [Lin and Cook, 2020]. With this capability, we can cluster participants based on these learned routines and thus obtain more nuanced subgroups for intervention policy learning and evaluation.

References

[Abbeel and Ng, 2004] Abbeel, P. and Ng, A. Y. (2004). Apprenticeship learning via inverse reinforcement learning. In *Proceedings of the twenty-first international conference on Machine learning*, page 1.

[Aldao, 2013] Aldao, A. (2013). The Future of Emotion Regulation Research: Capturing Context. *Perspectives on Psychological Science*, 8(2):155–172.

[Aldao and Nolen-Hoeksema, 2013] Aldao, A. and Nolen-Hoeksema, S. (2013). One versus many: Capturing the use of multiple emotion regulation strategies in response to an emotion-eliciting stimulus. *Cognition and Emotion*, 27(4):753–760.

[Aldao et al., 2010a] Aldao, A., Nolen-Hoeksema, S., and Schweizer, S. (2010a). Emotion-regulation strategies across psychopathology: A meta-analytic review. *Clinical Psychology Review*, 30(2):217 – 237.

[Aldao et al., 2010b] Aldao, A., Nolen-Hoeksema, S., and Schweizer, S. (2010b). Emotion-regulation strategies across psychopathology: A meta-analytic review. *Clinical psychology review*, 30(2):217–237.

[Ameko et al., 2020] Ameko, M. K., Beltzer, M. L., Cai, L., Boukhechba, M., Teachman, B. A., and Barnes, L. E. (2020). Offline contextual multi-armed bandits for mobile health interventions: A case study on emotion regulation. In *Fourteenth ACM Conference on Recommender Systems*, pages 249–258.

[Ameko et al., 2018] Ameko, M. K., Cai, L., Boukhechba, M., Daros, A., Chow, P. I., Teachman, B. A., Gerber, M. S., and Barnes, L. E. (2018). Cluster-based approach to improve affect recognition from passively sensed data. In *2018 IEEE EMBS international conference on biomedical & health informatics (BHI)*, pages 434–437. IEEE.

[America,] America, M. H. Mental Health In America - Access To Care Data.

[Arumugam et al., 2018] Arumugam, D., Abel, D., Asadi, K., Gopalan, N., Grimm, C., Lee, J. K., Lehnert, L., and Littman, M. L. (2018). Mitigating planner overfitting in model-based reinforcement learning. *arXiv preprint arXiv:1812.01129*.

[Asselbergs et al., 2016] Asselbergs, J., Ruwaard, J., Ejdys, M., Schrader, N., Sijbrandij, M., and Riper, H. (2016). Mobile phone-based unobtrusive ecological momentary assessment of day-to-day mood: an explorative study. *Journal of medical Internet research*, 18(3).

[Atan et al., 2018] Atan, O., Jordon, J., and van der Schaar, M. (2018). Deep-treat: Learning optimal personalized treatments from observational data using neural networks. In *Thirty-Second AAAI Conference on Artificial Intelligence*.

[Beltzer et al., 2020] Beltzer, M. L., Ameko, M. K., Daniel, K. E., Daros, A. R., Boukhechba, M., Barnes, L. E., and Teachman, B. A. (2020). Building an emotion regulation recommender algorithm for socially anxious individuals using contextual bandits. *British Journal of Clinical Psychology*.

[Berle et al., 2011] Berle, D., Starcevic, V., Moses, K., Hannan, A., Milicevic, D., and Sammut, P. (2011). Preliminary validation of an ultra-brief version of the penn state worry questionnaire. *Clinical psychology & psychotherapy*, 18(4):339–346.

[Beygelzimer and Langford, 2009] Beygelzimer, A. and Langford, J. (2009). The offset tree for learning with partial labels. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 129–138.

[Bonanno and Burton, 2013] Bonanno, G. A. and Burton, C. L. (2013). Regulatory Flexibility: An Individual Differences Perspective on Coping and Emotion Regulation. *Perspectives on Psychological Science*, 8(6):591–612.

[Bostwick and Pankratz, 2000] Bostwick, J. M. and Pankratz, V. S. (2000). Affective disorders and suicide risk: A reexamination. *American Journal of Psychiatry*, 157(12):1925–1932.

[Boukhechba et al., 2018] Boukhechba, M., Daros, A. R., Fua, K., Chow, P. I., Teachman, B. A., and Barnes, L. E. (2018). Demonicsalmon: Monitoring mental health and social interactions of college students using smartphones. *Smart Health*, 9:192–203.

[Chakraborty, 2013] Chakraborty, B. (2013). *Statistical methods for dynamic treatment regimes*. Springer.

[Cheung et al., 2018] Cheung, K., Ling, W., Karr, C. J., Weingardt, K., Schueller, S. M., and Mohr, D. C. (2018). Evaluation of a recommender app for apps for the treatment of depression and anxiety: an analysis of longitudinal user engagement. *Journal of the American Medical Informatics Association*, 25(8):955–962.

[Chipman et al., 2010] Chipman, H. A., George, E. I., McCulloch, R. E., et al. (2010). Bart: Bayesian additive regression trees. *The Annals of Applied Statistics*, 4(1):266–298.

[Chow et al., 2017] Chow, P. I., Fua, K., Huang, Y., Bonelli, W., Xiong, H., Barnes, L. E., and Teachman, B. A. (2017). Using mobile sensing to test clinical models of depression, social anxiety, state affect, and social isolation among college students. *Journal of medical Internet research*, 19(3).

[Clark and Watson, 1988] Clark, L. A. and Watson, D. (1988). Mood and the mundane: Relations between daily life events and self-reported mood. *Journal of personality and social psychology*, 54(2):296.

[Clark et al., 1994] Clark, L. A., Watson, D., and Mineka, S. (1994). Temperament, personality, and the mood and anxiety disorders. *Journal of Abnormal Psychology*, 103(1):103.

[Clifton et al., 2013] Clifton, L., Clifton, D. A., Pimentel, M. A., Watkinson, P. J., and Tarassenko, L. (2013). Gaussian processes for personalized e-health monitoring with wearable sensors. *IEEE Transactions on Biomedical Engineering*, 60(1):193–197.

[Daros et al., 2020] Daros, A. R., Daniel, K. E., Boukhechba, M., Chow, P. I., Barnes, L. E., and Teachman, B. A. (2020). Relationships between trait emotion dysregulation and emotional experiences in daily life: an experience sampling study. *Cognition and Emotion*, 34(4):743–755.

[Dixon-Gordon et al., 2015] Dixon-Gordon, K. L., Aldao, A., and De Los Reyes, A. (2015). Emotion regulation in context: Examining the spontaneous use of strategies across emotional intensity and type of emotion. *Personality and Individual Differences*, 86:271–276.

[Doherty et al., 2020] Doherty, K., Balaskas, A., and Doherty, G. (2020). The design of ecological momentary assessment technologies. *Interacting with Computers*, 32(3):257–278.

[Doré et al., 2016] Doré, B. P., Silvers, J. A., and Ochsner, K. N. (2016). Toward a Personalized Science of Emotion Regulation. *Social and Personality Psychology Compass*, 10(4):171–187.

[Dudík et al., 2014] Dudík, M., Erhan, D., Langford, J., Li, L., et al. (2014). Doubly robust policy evaluation and optimization. *Statistical Science*, 29(4):485–511.

[Dudik et al., 2011] Dudik, M., Hsu, D., Kale, S., Karampatziakis, N., Langford, J., Reyzin, L., and Zhang, T. (2011). Efficient optimal learning for contextual bandits. *arXiv preprint arXiv:1106.2369*.

[Egloff et al., 1995] Egloff, B., Tausch, A., Kohlmann, C.-W., and Krohne, H. W. (1995). Relationships between time of day, day of the week, and positive mood: Exploring the role of the mood measure. *Motivation and emotion*, 19(2):99–110.

[el Hassouni et al., 2018] el Hassouni, A., Hoogendoorn, M., van Otterlo, M., and Barbaro, E. (2018). Personalization of health interventions using cluster-based reinforcement learning. In *International Conference on Principles and Practice of Multi-Agent Systems*, pages 467–475. Springer.

[Fernandez et al., 2016] Fernandez, K. C., Jazaieri, H., and Gross, J. J. (2016). Emotion Regulation: A Transdiagnostic Perspective on a New RDoC Domain. *Cognitive Therapy and Research*, 40(3):426–440.

[Ford and Mauss, 2015] Ford, B. and Mauss, I. (2015). Culture and emotion regulation (vol. 3). *Current Opinion in Psychology*, pages 1–5.

[Forman et al., 2019] Forman, E. M., Kerrigan, S. G., Butryn, M. L., Juarascio, A. S., Manasse, S. M., Ontañón, S., Dallal, D. H., Crochiere, R. J., and Moskow, D. (2019). Can the artificial intelligence technique of reinforcement learning use continuously-monitored digital data to optimize treatment for weight loss? *Journal of behavioral medicine*, 42(2):276–290.

[François-Lavet et al., 2015] François-Lavet, V., Fonteneau, R., and Ernst, D. (2015). How to discount deep reinforcement learning: Towards new dynamic strategies. *arXiv preprint arXiv:1512.02011*.

[Freitas and Salovey, 2000] Freitas, A. and Salovey, P. (2000). Regulating emotion in the short and long term. *Psychological Inquiry*, 11(3):178–179.

[Fu et al., 2020] Fu, J., Kumar, A., Nachum, O., Tucker, G., and Levine, S. (2020). D4rl: Datasets for deep data-driven reinforcement learning. *arXiv preprint arXiv:2004.07219*.

[Gentzler and Kerns, 2006] Gentzler, A. and Kerns, K. (2006). Adult attachment and memory of emotional reactions to negative and positive events. *Cognition & Emotion*, 20(1):20–42.

[Goldenberg et al., 2021] Goldenberg, D., Kofman, K., Albert, J., Mizrachi, S., Horowitz, A., and Teinemaa, I. (2021). Personalization in practice: Methods and applications. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, pages 1123–1126.

[Grant et al., 2005] Grant, B. F., Hasin, D. S., Blanco, C., Stinson, F. S., Chou, S. P., Goldstein, R. B., Dawson, D. A., Smith, S., Saha, T. D., and Huang, B. (2005). The epidemiology of social anxiety disorder in the united states: results from the national epidemiologic survey on alcohol and related conditions. *Journal of Clinical Psychiatry*, 66(11):1351–1361.

[Gross, 1998] Gross, J. J. (1998). The emerging field of emotion regulation: An integrative review. *Review of General Psychology*, 2(3):271–299.

[Gross and John, 2003] Gross, J. J. and John, O. P. (2003). Individual differences in two emotion regulation processes: implications for affect, relationships, and well-being. *Journal of personality and social psychology*, 85(2):348.

[Gulcehre et al., 2020] Gulcehre, C., Wang, Z., Novikov, A., Paine, T. L., Colmenarejo, S. G., Zolna, K., Agarwal, R., Merel, J., Mankowitz, D., Paduraru, C., et al. (2020). Rl unplugged: Benchmarks for offline reinforcement learning. *arXiv preprint arXiv:2006.13888*.

[Guszkowska, 2004] Guszkowska, M. (2004). Effects of exercise on anxiety, depression and mood. *Psychiatria polska*, 38(4):611–620.

[Hachiya et al., 2012] Hachiya, H., Sugiyama, M., and Ueda, N. (2012). Importance-weighted least-squares probabilistic classifier for covariate shift adaptation with application to human activity recognition. *Neurocomputing*, 80:93–101.

[Hamerly and Elkan, 2004] Hamerly, G. and Elkan, C. (2004). Learning the k in k-means. In *Advances in neural information processing systems*, pages 281–288.

[Heimberg et al., 1992] Heimberg, R. G., Mueller, G. P., Holt, C. S., Hope, D. A., and Liebowitz, M. R. (1992). Assessment of anxiety in social interaction and being observed by others: The social interaction anxiety scale and the social phobia scale. *Behavior therapy*, 23(1):53–73.

[Heiy and Cheavens, 2014] Heiy, J. E. and Cheavens, J. S. (2014). Back to Basics: A Naturalistic Assessment of the Experience and Regulation of Emotion. *Emotion*, 14(5):878–891.

[Hilty et al., 2013] Hilty, D. M., Ferrer, D. C., Parish, M. B., Johnston, B., Callahan, E. J., and Yellowlees, P. M. (2013). The effectiveness of telemental health: A 2013 review. *Telemedicine and e-Health*, 19(6):444–454.

[Huang, 1998] Huang, Z. (1998). Extensions to the k-means algorithm for clustering large data sets with categorical values. *Data mining and knowledge discovery*, 2(3):283–304.

[Imbens and Rubin, 2015] Imbens, G. W. and Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

[Jaimes et al., 2014] Jaimes, L. G., Llofriu, M., and Raij, A. (2014). A stress-free life: just-in-time interventions for stress via real-time forecasting and intervention adaptation. In *Proceedings of the 9th International Conference on Body Area Networks*, pages 197–203. ICST (Institute for Computer Sciences, Social-Informatics and ....

[Jaques et al., 2016] Jaques, N., Taylor, S., Nosakhare, E., Sano, A., and Picard, R. (2016). Multi-task learning for predicting health, stress, and happiness. In *NIPS Workshop on Machine Learning for Healthcare*.

[Jaques et al., 2017] Jaques, N., Taylor, S., Sano, A., Picard, R., et al. (2017). Predicting tomorrow's mood, health, and stress level using personalized multitask learning and domain adaptation. In *IJCAI 2017 Workshop on artificial intelligence in affective computing*, pages 17–33.

[Jiang et al., 2015] Jiang, N., Kulesza, A., Singh, S., and Lewis, R. (2015). The dependence of effective planning horizon on model accuracy. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*, pages 1181–1189. Citeseer.

[Jiang and Li, 2016] Jiang, N. and Li, L. (2016). Doubly robust off-policy value evaluation for reinforcement learning. In *International Conference on Machine Learning*, pages 652–661. PMLR.

[Kang et al., 2007] Kang, J. D., Schafer, J. L., et al. (2007). Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical science*, 22(4):523–539.

[Kang et al., 2005] Kang, J. H., Welbourne, W., Stewart, B., and Borriello, G. (2005). Extracting places from traces of locations. *ACM SIGMOBILE Mobile Computing and Communications Review*, 9(3):58.

[Kaufman et al., 2016] Kaufman, E. A., Xia, M., Fosco, G., Yaptangco, M., Skidmore, C. R., and Crowell, S. E. (2016). The difficulties in emotion regulation scale short form (ders-sf): Validation and replication in adolescent and adult samples. *Journal of Psychopathology and Behavioral Assessment*, 38(3):443–455.

[Kessler et al., 2005a] Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., and Walters, E. E. (2005a). Lifetime prevalence and age-of-onset distributions of dsm-iv disorders in the national comorbidity survey replication. *Archives of General Psychiatry*, 62(6):593–602.

[Kessler et al., 2005b] Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., and Walters, E. E. (2005b). Lifetime Prevalence and Age-of-Onset Distributions of DSM-IV Disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry*, 62(June):593–602.

[Keßler, 2015] Keßler, C. (2015). OpenStreetMap. In Shekhar, S., Xiong, H., and Zhou, X., editors, *Encyclopedia of GIS*, pages 1–5. Springer International Publishing, Cham. DOI: 10.1007/978-3-319-23519-6_1654-1.

[Klasnja et al., 2015] Klasnja, P., Hekler, E. B., Shiffman, S., Boruvka, A., Almirall, D., Tewari, A., and Murphy, S. A. (2015). Microrandomized trials: An experimental design for developing just-in-time adaptive interventions. *Health Psychology*, 34(S):1220.

[Koldijk et al., 2016] Koldijk, S., Neerincx, M. A., and Kraaij, W. (2016). Detecting work stress in offices by combining unobtrusive sensors. *IEEE Transactions on affective computing*, 9(2):227–239.

[Komorowski et al., 2018] Komorowski, M., Celi, L. A., Badawi, O., Gordon, A. C., and Faisal, A. A. (2018). The artificial intelligence clinician learns optimal treatment strategies for sepsis in intensive care. *Nature medicine*, 24(11):1716–1720.

[Kroenke et al., 2009] Kroenke, K., Strine, T. W., Spitzer, R. L., Williams, J. B., Berry, J. T., and Mokdad, A. H. (2009). The phq-8 as a measure of current depression in the general population. *Journal of affective disorders*, 114(1-3):163–173.

[Li et al., 2010] Li, L., Chu, W., Langford, J., and Schapire, R. E. (2010). A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th international conference on World wide web*, pages 661–670.

[Liao et al., 2020a] Liao, P., Greenewald, K., Klasnja, P., and Murphy, S. (2020a). Personalized heartsteps: A reinforcement learning algorithm for optimizing physical activity. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4(1):1–22.

[Liao et al., 2020b] Liao, P., Klasnja, P., and Murphy, S. (2020b). Off-policy estimation of long-term average outcomes with applications to mobile health. *Journal of the American Statistical Association*, pages 1–10.

[Liao et al., 2020c] Liao, P., Qi, Z., and Murphy, S. (2020c). Batch policy learning in average reward markov decision processes. *arXiv preprint arXiv:2007.11771*.

[LiKamWa et al., 2013] LiKamWa, R., Liu, Y., Lane, N. D., and Zhong, L. (2013). Moodscope: Building a mood sensor from smartphone usage patterns. In *Proceeding of the 11th annual international conference on Mobile systems, applications, and services*, pages 389–402. ACM.

[Lin and Cook, 2020] Lin, B. and Cook, D. J. (2020). Analyzing sensor-based individual and population behavior patterns via inverse reinforcement learning. *Sensors*, 20(18):5207.

[Lin et al., 2018] Lin, L., Stamm, K., and Christidis, P. (2018). Demographics of the U.S. Psychology Workforce: Findings from the 2007-16 American Community Survey. Technical Report May, American Psychological Association Center for Workforce Studies.

[Marinier et al., 2008] Marinier, R. P., Laird, J. E., and Marinier Iii, R. P. (2008). Emotion-Driven Reinforcement Learning. *Proceedings of the 30th Annual Conference of the Cognitive Science Society*, pages 115–120.

[Mattick and Clarke, 1998] Mattick, R. P. and Clarke, J. C. (1998). Development and validation of measures of social phobia scrutiny fear and social interaction anxiety. *Behaviour research and therapy*, 36(4):455–470.

[McCaffrey et al., 2013] McCaffrey, D. F., Griffin, B. A., Almirall, D., Slaughter, M. E., Ramchand, R., and Burgette, L. F. (2013). A tutorial on propensity score estimation for multiple treatments using generalized boosted models. *Statistics in medicine*, 32(19):3388–3414.

[McInerney et al., 2021] McInerney, J., Elahi, E., Basilico, J., Raimond, Y., and Jebara, T. (2021). Accordion: A trainable simulator forlong-term interactive systems. In *Fifteenth ACM Conference*

on Recommender Systems, RecSys '21, page 102–113, New York, NY, USA. Association for Computing Machinery.

[Morgan and Winship, 2015] Morgan, S. L. and Winship, C. (2015). *Counterfactuals and causal inference*. Cambridge University Press.

[Murray et al., 2013] Murray, C. J., Abraham, J., Ali, M. K., Alvarado, M., Atkinson, C., Baddour, L. M., Bartels, D. H., Benjamin, E. J., Bhalla, K., Birbeck, G., et al. (2013). The state of us health, 1990-2010: burden of diseases, injuries, and risk factors. *Jama*, 310(6):591–606.

[Nahum-Shani et al., 2017] Nahum-Shani, I., Smith, S. N., Spring, B. J., Collins, L. M., Witkiewitz, K., Tewari, A., and Murphy, S. A. (2017). Just-in-time adaptive interventions (jitais) in mobile health: key components and design principles for ongoing health behavior support. *Annals of Behavioral Medicine*, 52(6):446–462.

[Nolen-Hoeksema and Aldao, 2011a] Nolen-Hoeksema, S. and Aldao, A. (2011a). Gender and age differences in emotion regulation strategies and their relationship to depressive symptoms. *Personality and Individual Differences*, 51(6):704–708.

[Nolen-Hoeksema and Aldao, 2011b] Nolen-Hoeksema, S. and Aldao, A. (2011b). Gender and age differences in emotion regulation strategies and their relationship to depressive symptoms. *Personality and individual differences*, 51(6):704–708.

[Paredes et al., 2014] Paredes, P., Gilad-Bachrach, R., Czerwinski, M., Roseway, A., Rowan, K., and Hernandez, J. (2014). Poptherapy: Coping with stress through pop-culture. In *Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare*, pages 109–117. ICST (Institute for Computer Sciences, Social-Informatics and ....

[Pearl, 2009] Pearl, J. (2009). *Causality*. Cambridge university press.

[Peng et al., 2018] Peng, X., Ding, Y., Wihl, D., Gottesman, O., Komorowski, M., Lehman, L.-w. H., Ross, A., Faisal, A., and Doshi-Velez, F. (2018). Improving sepsis treatment strategies by combining deep and kernel-based reinforcement learning. In *AMIA Annual Symposium Proceedings*, volume 2018, page 887. American Medical Informatics Association.

[Rabbi et al., 2017] Rabbi, M., Aung, M. H., and Choudhury, T. (2017). Towards health recommendation systems: an approach for providing automated personalized health feedback from mobile data. In *Mobile Health*, pages 519–542. Springer.

[Rabbi et al., 2015] Rabbi, M., Aung, M. H., Zhang, M., and Choudhury, T. (2015). Mybehavior: automatic personalized health feedback from user behaviors and preferences using smartphones. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 707–718. ACM.

[Raghu et al., 2017] Raghu, A., Komorowski, M., Celi, L. A., Szolovits, P., and Ghassemi, M. (2017). Continuous state-space models for optimal sepsis treatment-a deep reinforcement learning approach. *arXiv preprint arXiv:1705.08422*.

[Rahman et al., 2016] Rahman, T., Czerwinski, M., Gilad-Bachrach, R., and Johns, P. (2016). Predicting about-to-eat moments for just-in-time eating intervention. In *Proceedings of the 6th International Conference on Digital Health Conference*, pages 141–150. ACM.

[Raio et al., 2016] Raio, C. M., Goldfarb, E. V., Lempert, K. M., and Sokol-Hessner, P. (2016). Classifying emotion regulation strategies. *Nature Reviews Neuroscience*, 17(8):532.

[Rakelly et al., 2019] Rakelly, K., Zhou, A., Finn, C., Levine, S., and Quillen, D. (2019). Efficient off-policy meta-reinforcement learning via probabilistic context variables. In *International conference on machine learning*, pages 5331–5340. PMLR.

[Rashid et al., 2020] Rashid, H., Mendu, S., Daniel, K. E., Beltzer, M. L., Teachman, B. A., Boukhechba, M., and Barnes, L. E. (2020). Predicting subjective measures of social anxiety from sparsely collected mobile sensor data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4(3):1–24.

[Rasmussen and Williams,] Rasmussen, C. E. and Williams, C. K. *Gaussian processes for machine learning*, volume 1.

[Ridgeway et al.,] Ridgeway, G., McCaffrey, D., Morral, A., Burgette, L., and Griffin, B. A. Toolkit for weighting and analysis of nonequivalent groups: A tutorial for the twang package.

[Rodebaugh et al., 2007] Rodebaugh, T. L., Woods, C. M., and Heimberg, R. G. (2007). The reverse of social anxiety is not always the opposite: The reverse-scored items of the social interaction anxiety scale do not belong. *Behavior Therapy*, 38(2):192–206.

[Rohde et al., 2018] Rohde, D., Bonner, S., Dunlop, T., Vasile, F., and Karatzoglou, A. (2018). Recogym: A reinforcement learning environment for the problem of product recommendation in online advertising. *arXiv preprint arXiv:1808.00720*.

[Rosenbaum and Rubin, 1983] Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55.

[Rousseeuw, 1987] Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65.

[Rubin, 1974] Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688.

[Saeb et al., 2015] Saeb, S., Zhang, M., Karr, C. J., Schueller, S. M., Corden, M. E., Kording, K. P., and Mohr, D. C. (2015). Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: An exploratory study. *Journal of Medical Internet Research*, 17(7):e175.

[Sheppes et al., 2014] Sheppes, G., Scheibe, S., Suri, G., Radu, P., Blechert, J., and Gross, J. J. (2014). Emotion regulation choice: A conceptual framework and supporting evidence. *Journal of Experimental Psychology: General*, 143(1):163–181.

[Shiffman et al., 2008] Shiffman, S., Stone, A. A., and Hufford, M. R. (2008). Ecological momentary assessment. *Annu. Rev. Clin. Psychol.*, 4:1–32.

[Sloan et al., 2017] Sloan, E., Hall, K., Moulding, R., Bryce, S., Mildred, H., and Staiger, P. K. (2017). Emotion regulation as a transdiagnostic treatment construct across anxiety, depression, substance, eating and borderline personality disorders: A systematic review. *Clinical Psychology Review*, 57(October 2016):141–163.

[Steger and Kashdan, 2009] Steger, M. F. and Kashdan, T. B. (2009). Depression and everyday social activity, belonging, and well-being. *Journal of counseling psychology*, 56(2):289.

[Strehl et al., 2010] Strehl, A., Langford, J., Li, L., and Kakade, S. M. (2010). Learning from logged implicit exploration data. In *Advances in Neural Information Processing Systems*, pages 2217–2225.

[Stuart et al., 2013] Stuart, E. A., Lee, B. K., and Leacy, F. P. (2013). Prognostic score–based balance measures can be a useful diagnostic for propensity score methods in comparative effectiveness research. *Journal of clinical epidemiology*, 66(8):S84–S90.

[Suri et al., 2018] Suri, G., Sheppes, G., Young, G., Abraham, D., McRae, K., and Gross, J. J. (2018). Emotion regulation choice: the role of environmental affordances. *Cognition and Emotion*, 32(5):963–971.

[Swaminathan and Joachims, 2015] Swaminathan, A. and Joachims, T. (2015). Batch learning from logged bandit feedback through counterfactual risk minimization. *The Journal of Machine Learning Research*, 16(1):1731–1755.

[Tekin et al., 2014] Tekin, C., Atan, O., and Van Der Schaar, M. (2014). Discover the expert: Context-adaptive expert selection for medical diagnosis. *IEEE transactions on emerging topics in computing*, 3(2):220–234.

[Tewari and Murphy, 2017] Tewari, A. and Murphy, S. A. (2017). From ads to interventions: Contextual bandits in mobile health. In *Mobile Health*, pages 495–517. Springer.

[Tomkins et al., 2021] Tomkins, S., Liao, P., Klasnja, P., and Murphy, S. (2021). Intelligentpooling: practical thompson sampling for mhealth. *Machine Learning*, pages 1–43.

[Troy et al., 2013] Troy, A. S., Shallcross, A. J., and Mauss, I. B. (2013). A Person-by-Situation Approach to Emotion Regulation: Cognitive Reappraisal Can Either Help or Hurt, Depending on the Context. *Psychological Science*, 24(12):2505–2514.

[Van Hasselt et al., 2015] Van Hasselt, H., Guez, A., and Silver, D. (2015). Deep reinforcement learning with double q-learning. *arXiv preprint arXiv:1509.06461*.

[Wager and Athey, 2018] Wager, S. and Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523):1228–1242.

[Wang et al., 2017] Wang, R., Chen, F., Chen, Z., Li, T., Harari, G., Tignor, S., Zhou, X., Ben-Zeev, D., and Campbell, A. T. (2017). StudentLife: Using smartphones to assess mental health and academic performance of college students. In *Mobile Health*, pages 7–33. Springer International Publishing.

[Watson et al., 1992] Watson, D., Clark, L. A., McIntyre, C. W., and Hamaker, S. (1992). Affect, personality, and social activity. *Journal of personality and social psychology*, 63(6):1011.

[Xiong et al., 2016] Xiong, H., Huang, Y., Barnes, L. E., and Gerber, M. S. (2016). Sensus: A cross-platform, general-purpose system for mobile crowdsensing in human-subject studies. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '16*. ACM Press.

[Xu et al., 2021] Xu, X., Chikersal, P., Dutcher, J. M., Sefidgar, Y. S., Seo, W., Tumminia, M. J., Villalba, D. K., Cohen, S., Creswell, K. G., Creswell, J. D., et al. (2021). Leveraging collaborative-filtering for personalized behavior modeling: A case study of depression detection among college students. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 5(1):1–27.

[Yang et al., 2018] Yang, S., Zhou, P., Duan, K., Hossain, M. S., and Alhamid, M. F. (2018). emhealth: towards emotion health through depression prediction and intelligent health recommender system. *Mobile Networks and Applications*, 23(2):216–226.

[Yom-Tov et al., 2017] Yom-Tov, E., Feraru, G., Kozdoba, M., Mannor, S., Tennenholtz, M., and Hochberg, I. (2017). Encouraging physical activity in patients with diabetes: intervention using a reinforcement learning system. *Journal of medical Internet research*, 19(10):e338.

[Yonekura et al., 2016] Yonekura, S., Okamura, S., Kajiwara, Y., and Shimakawa, H. (2016). Mood prediction reflecting emotion state to improve mental health. volume 3, pages 404–407.

[Zhang and Bareinboim, 2017] Zhang, J. and Bareinboim, E. (2017). Transfer learning in multi-armed bandit: a causal approach. In *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*, pages 1778–1780.

[Zhu and Liao, 2017] Zhu, F. and Liao, P. (2017). Effective warm start for the online actor-critic reinforcement learning based mhealth intervention. *arXiv preprint arXiv:1704.04866*.

Appendices

A | Appendix

A.1 SAMMI Data Streams

The SAMMI study was conducted over a period of 5 weeks for each participant, but there is a varied amount of data quality collected as well as study engagement observed. In this appendix, we present the plot of various data streams from the SAMMI study over time for a given subject ID and mobile device operating system.

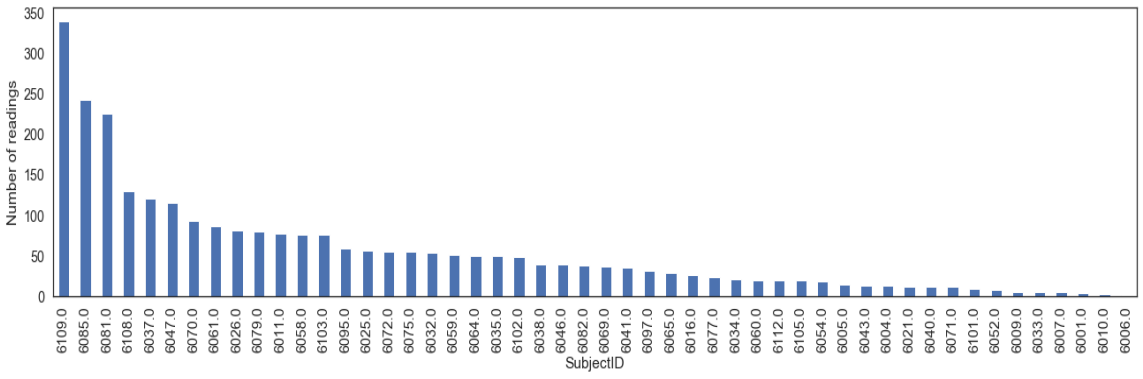


Figure A.1: Study engagement level for 50 sampled users.

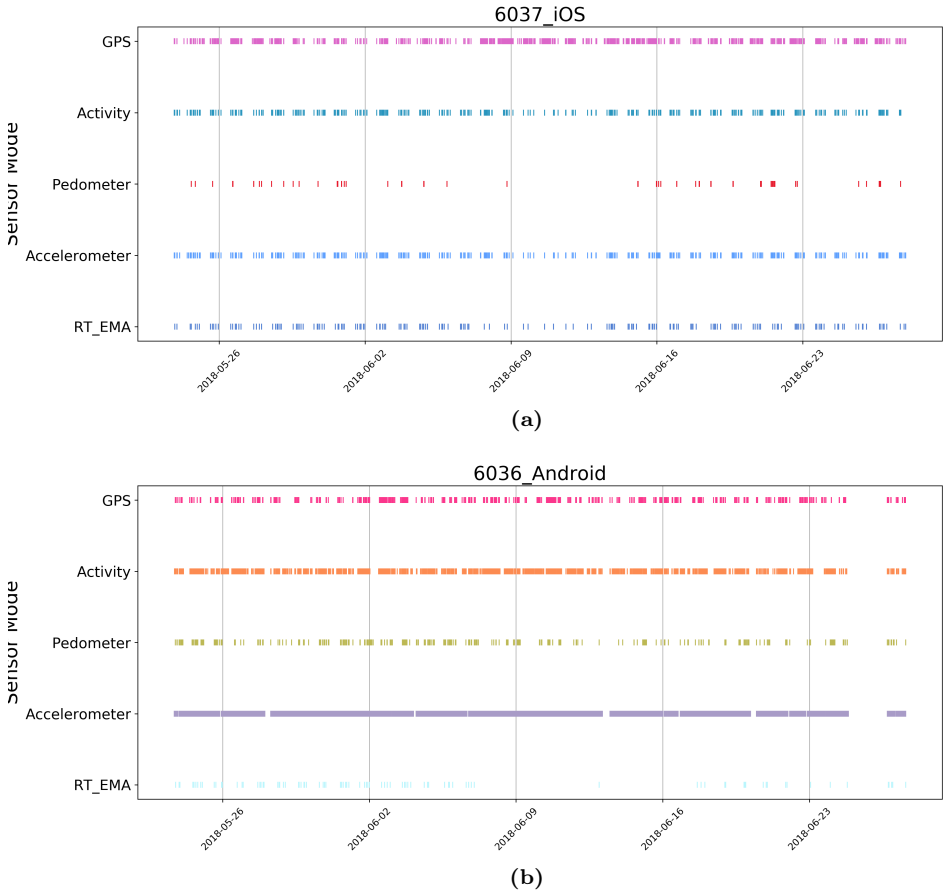


Figure A.2: Example of Subject IDs with relatively full data streams over the study period. Note that RT-EMA refers to randomly timed EMA surveys.

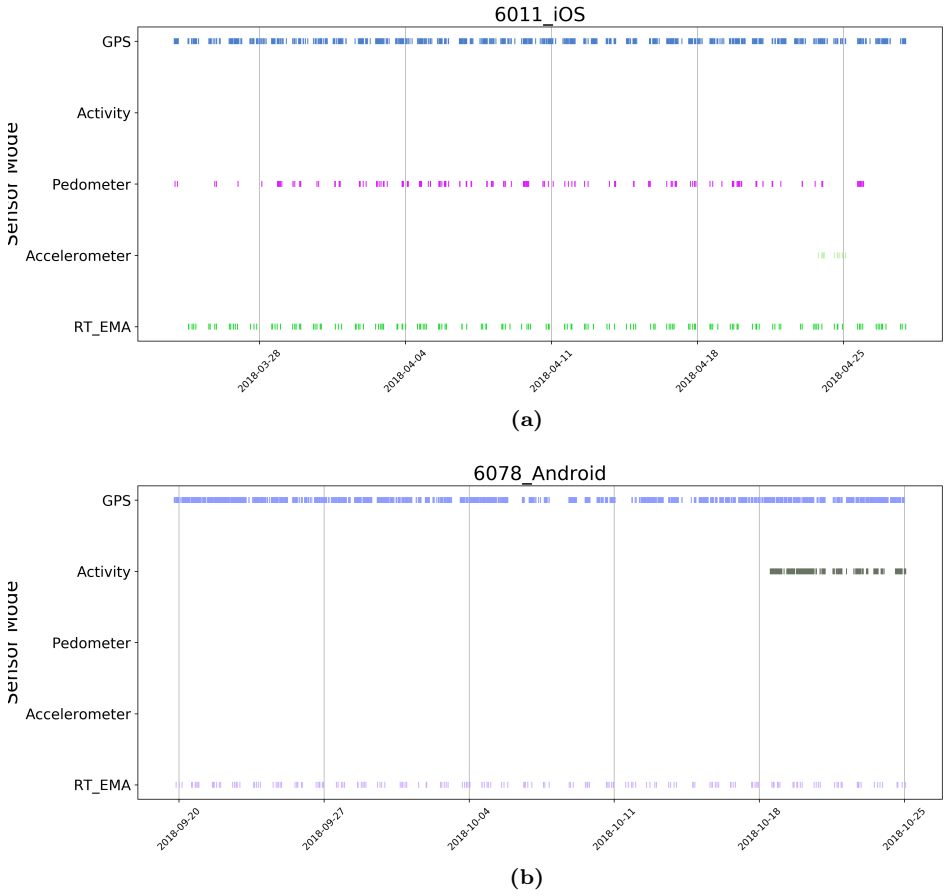


Figure A.3: Example of Subject IDs with relatively sparse data streams over the study period. Note that RT-EMA refers to randomly timed EMA surveys