Machine Learning Methods for Targeting and Personalization

Omid Rafieian[*]
Cornell Tech and Cornell University

Hema Yoganarasimhan[*]
University of Washington

Abstract

In this chapter, we review the methods available to effectively personalize marketing interventions in static settings. We focus on the recent methodological developments in this space, that combine ideas from machine learning and causal inference to automate the task of personalization. We present a series of approaches for this task and discuss their pros and cons. Specifically, we summarize three broad sets of methods – (1) Outcome modeling-based approaches, (2) Treatment effects-based approaches, and (3) Direct policy learning-based approaches. We also briefly list the various applied papers that use different methods. We then discuss the three evaluation approaches available to evaluate the gains from adopting personalized policies – the Direct method, the Inverse Propensity Score estimator, and the Doubly Robust method. We conclude with a discussion of special cases that go beyond the standard static setting. We conclude with some challenges and some directions for future research.

Keywords: Personalization, Policy Design, Causal Inference, Targeting, Machine Learning

[*] Please address all correspondence to: or83@cornell.edu, hemay@uw.edu.

1 Introduction

Machine learning methods have become a mainstream practice in marketing. Broadly, firms apply machine learning methods to three types of tasks: (1) descriptive, (2) predictive, and (3) prescriptive. Many marketing problems are descriptive in nature. Descriptive tasks often involve describing important aspects of large-scale structured or unstructured data. For example, a fashion brand may be interested in analyzing large-scale user-generated textual and visual content online to better understand how the brand is perceived by its customers (Liu et al., 2020). A suite of machine learning methods help with these tasks, ranging from simple clustering algorithms to more advanced topic modeling and representation learning algorithms that effectively process terabytes of unstructured inputs.

The second class of marketing tasks is predictive. In these tasks, we want to predict a certain outcome using a high-dimensional set of features. For example, an online retailer wants to predict its demand for the next month, given all the information available (e.g., past demand, user-generated content). Firms can use a vast array of algorithms to perform this task, including gradient-boosting algorithms and deep neural networks. Dzyabura and Yoganarasimhan (2018) provide a detailed summary of supervised learning algorithms that help with predictive tasks.

The final class of tasks is prescriptive, wherein the firm wants to use machine learning for decision-making. For example, a mobile app may want to deliver personalized notifications to each user based on their individual characteristics to maximize user engagement. Machine learning methods for this task generally involve learning a policy from the interactions in the data and are often referred to as reinforcement learning (Sutton and Barto, 2018). In this chapter, we focus on a prescriptive task of long-standing interest in the field of marketing: personalization (Rossi et al., 1996).

Personalization at scale has long been the ultimate goal of marketers, captivating the interest of both academics and industry professionals. Yet, achieving individual-level personalization on a large scale was, until recently, more of a theoretical concept than a practical reality. However, with advancements in computing power, data storage, and the convergence of machine learning with causal inference, this elusive goal is now within grasp. In recent years, interdisciplinary research spanning marketing, computer science, economics, and statistics has tackled this challenge from multiple angles, providing solutions that are not only theoretically grounded but also operationally viable. These innovative approaches have found application across various sectors, from tailoring news and content recommendations online to crafting personalized promotions in retail environments. As these applications have proliferated, our knowledge and understanding of this area has grown significantly.

This chapter reviews the recently developed machine learning methods available for personalization in static settings. That is, we focus on how a firm or policy maker can personalize their marketing interventions (e.g., product, price, advertising, promotions) to optimize some outcome measure of interest (e.g., consumer acquisition, retention, profit, revenue, clicks/conversions). In §2, we first present the general problem set-up, the notation used, and a mathematical definition of the personalization problem. These notations and definitions will form the foundation on which all the solution concepts will be built. Next, in §3–§5, we discuss the three broad sets of methods available for personalized policy design – (1) Outcome modeling-based approaches, (2) Treatment effects-based approaches, and (3) Direct policy learning-based approaches. We also explain the pros and cons of each and discuss how we need to combine ideas from causal inference and machine learning to effectively design policies that have counterfactual validity. In §6, we present the approaches available to evaluate personalized policies and cover some best practices. Next, §7, we briefly discuss the approaches used in cases beyond the standard static one-shot personalization problems. Finally, in §8, we conclude with a discussion of the remaining challenges and some directions for future research.

In this review, we focus specifically on methodological approaches for developing and evaluating personalized targeting policies in static settings. As such, we do not provide a comprehensive summary of the application papers in this area. We refer interested readers to Rafieian and Yoganarasimhan (2023) for a discussion of the substantive findings on personalization.[1]

2 General Model Setup

Our goal in this section is to introduce a unified framework that can flexibly model a wide range of personalization problems. At a high level, personalization happens when a firm or social planner (henceforth referred to as the planner) differentiates between individual units of the population by assigning them to different treatments. As such, for personalization to be possible, the planner must be able to (1) differentiate between units of population and (2) control the treatment assignment. Data collection at the individual level enables the former, whereas building an infrastructure to deliver the treatment at the individual level enables the latter. The main goal of this exercise is to achieve better outcomes/rewards of interest to the planner.

Consider a case where there are N individuals indexed by subscript i. As shown in Figure 1, we can assume a general timeline for the personalization problem that involves three steps:

1. In the first step, the planner observes an individual i from the population with a set of

[1] Interested readers can also refer to Murthi and Sarkar (2003), Steckel et al. (2005), Proserpio et al. (2020), and Liaukonyte (2021) for earlier summaries of personalization across a wide variety of contexts.

characteristics $X_i \in \mathcal{X}$. Individual-level characteristics are often referred to as covariates, pre-treatment variables, or the context vector (and the exact nomenclature varies depending on the stream of the literature). The richness of individual-level characteristics determines the planner's ability to differentiate between population units. As a result, planners generally like to collect massive amounts of individual-level data (e.g., technology firms and digital platforms). Since collecting individual-level data undermines individuals' privacy, data protection policies are often designed to limit the richness and accuracy of X.

2. In the second step, the planner decides which treatment to deliver to a given individual. We denote the treatment variable for the i-th individual by $W_i \in \mathcal{W}$. The planner can fully control the assignment to the treatment variable W_i. We define the set of treatments as $\mathcal{W} = \{w^{(0)}, w^{(1)}, \dots, w^{(J-1)}\}$. We do not impose any restriction on the set $\mathcal{W}$, as it can be very large (e.g., ad copies) and/or multi-dimensional (e.g., products with different prices and promotions).

3. In the third step, the individual-level outcomes $Y_i \in \mathcal{Y}$ are realized depending on the individual's set of characteristics X and treatment assignment W. We let $\{Y_i(w)\}_{w \in \mathcal{W}}$ denote the set of potential outcomes, such that $Y_i = Y_i(W_i)$. For notation simplicity, we use shorthand $Y(j) \coloneqq Y(w^{(j)})$. The outcomes are often referred to as rewards in the personalization literature. We define an expected reward function as $R : \mathcal{X} \times \mathcal{W} \to \mathcal{Y}$, which determines the expected outcome given the set of covariates and treatment assignment as follows:

$$R(x, w) = \mathbb{E}[Y(w) \mid X = x] \tag{1}$$

We set treatment $w^{(0)}$ as the baseline and define conditional average treatment effect for treatment $w^{(j)}$ as follows:

$$\tau_w(x) = \mathbb{E}[Y(w) - Y(0) \mid X = x], \tag{2}$$

where $\tau_w(\cdot)$ is the conditional average treatment effect (CATE) function and can be any general function. We can now present the potential outcomes in the generic form below:

$$Y_i(w) = R(X_i, 0) + \tau_w(X_i) + \epsilon_i(w), \tag{3}$$

where $\epsilon_i(w)$ is the randomness in the process. In business applications, the outcome or reward is often closely tied to the planner's utility (e.g., ad revenue, sales).

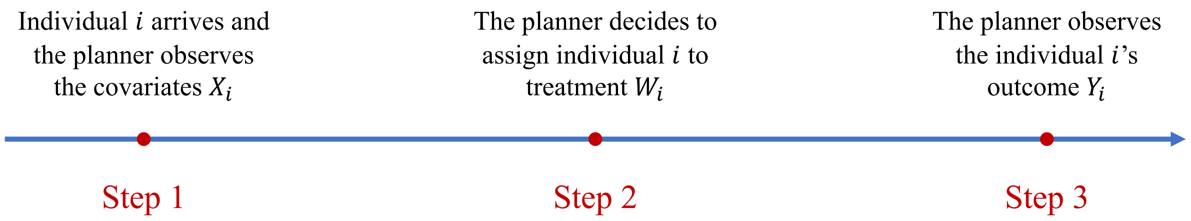
Individual i arrives and the planner observes the covariates X_i
The planner decides to assign individual i to treatment W_i
The planner observes the individual i's outcome Y_i
Step 1
Step 2
Step 3

Figure 1: Timeline of personalization

These three steps constitute the data set $\mathcal{D} = \{(X_i, W_i, Y_i)\}_{i=1}^{N}$ available to researchers, where i indexes the individuals in the data, and X_i, W_i, and Y_i denote the covariates, treatment assignment, and outcomes for individual i, respectively.

With all the model preliminaries defined, we now formally define a personalized policy as follows:

Definition 1. *A personalized policy is a mapping $\pi : \mathcal{W} \times \mathcal{X} \rightarrow [0, 1]$ that determines the probability for each treatment $w \in \mathcal{W}$ given that the individual's set of characteristics X. We denote the probability of action w for the individual with characteristics x as $\pi(w \mid x)$.*

This definition of personalized policy allows for both deterministic and probabilistic policies. For example, if $\pi(w \mid x) = 1$, it means that treatment w will deterministically be assigned to the individual with the set of characteristics X_i. However, if $\pi(w^{(1)} \mid x) = 0.7$ and $\pi(w^{(2)} \mid x) = 0.3$, it means that treatment $w^{(1)}$ will be delivered with probability 0.7 and treatment $w^{(2)}$ will be delivered with probability 0.3. Given the definition of a personalized policy in Definition 1, we can define the personalized policy that generated the data $\mathcal{D}$ as follows:

Definition 2. *The policy $\pi_{\mathcal{D}}(\cdot \mid \cdot)$ characterizes the policy that assigns treatments to individuals in the data. That is, $\pi_{\mathcal{D}}(w \mid X_i) = \Pr(W_i = w \mid X_i)$.*

In the causal inference literature, $\pi_{\mathcal{D}}(W_i \mid X_i)$ determines what is often referred to as the propensity score (Rosenbaum and Rubin, 1983). Knowledge of this policy function is important in identifying and evaluating personalized policies.

Finally, we define the performance or reward function that maps a policy to its performance. We denote the policy space by Π and define the performance function as follows:

Definition 3. *For an outcome Y, the performance function $\rho : \Pi \rightarrow \mathcal{Y}$ evaluates the performance of each policy in terms of the expected outcome as follows:*

$$\rho(\pi) = \mathbb{E}\left[\sum_{w \in \mathcal{W}} \pi(w \mid X_i) Y_i(w)\right] \tag{4}$$

In this chapter, we consider personalization in static settings with no interference, i.e., situations where the planner's decision to assign an individual to treatment only affects the reward in the current period, and does not affect future rewards or other individuals. With the definitions above, we can now define the planner's personalization problem as follows:

Definition 4. *Suppose there is a planner who assigns individuals to treatments in a static environment. The planner's decision for individual i has no consequence beyond determining the reward for individual i. For an individual of the set of characteristics x, the planner has the following objective function:*

$$\pi^* = \operatorname*{argmax}_{\pi} \rho(\pi) \tag{5}$$

To find the solution to the optimization problem in Equation (6), the planner can use a variety of methods. In this chapter, we discuss three solution concepts that differ in the way they view the objective function. We first present the outcome modeling approach in §3. Next, in §4, we present the solution based on causal modeling and discuss the causal nature of the personalization problem. Finally, in §5, we present another solution concept whereby the planner directly optimizes the policy. Our goal is to present both the advantages and disadvantages of each method, thereby providing a set of guidelines for researchers on which method(s) are most suitable in their settings.

3 Outcome Modeling

The first set of methods that aim to find a solution to the optimization problem in Equation (6) focus on outcome estimation. These methods view the optimization problem in Equation (5) as follows:

$$\operatorname*{argmax}_{\pi} \mathbb{E}\left[\sum_{w \in \mathcal{W}} \pi(w \mid X_i) R(X_i, w)\right], \tag{6}$$

where the expectation is taken over the randomness in the reward and policy functions. As such, these methods first obtain an estimate $\hat{R}$ of the expected reward function and then use this estimate to find the policy that performs best. We can summarize the steps in this approach as follows:

- **Step 1:** We first use the data at hand $\mathcal{D} = \{(X_i, W_i, Y_i)\}_i$ to estimate the function $R(x, w)$ as accurately as possible. We denote the estimate by $\hat{R}(\cdot, \cdot)$.
- **Step 2:** We then use the expected reward estimates to identify the treatment that maximizes the estimated expected rewards as follows:

$$\pi(w \mid x) = 1 \quad \text{if and only if} \quad w = \operatorname*{argmax}_{\tilde{w} \in \mathcal{W}} \hat{R}(x, \tilde{w}) \tag{7}$$

The performance of personalized policies developed under the outcome modeling approach depends heavily on how well we can estimate the outcomes. If we can perfectly estimate the expected reward function R, we can show that the outcome modeling approach finds the true optimal personalized policy. However, in reality, we cannot perfectly estimate R and our estimates likely have some error. Thus, the question of which method to use for a certain outcome modeling task is greatly important. In §3.1, we consider the case where the planner has a rich set of user-level covariates and present the supervised learning solution to this problem. Next, in §3.2, we focus on the problem with a large treatment space and turn to low-rank methods as the solution.

3.1 Supervised Learning with Rich Covariates

Suppose the planner has a training data set $\mathcal{D}_{\text{train}} = \{(X_i, W_i, Y_i)\}_{i=1}^{N}$, where the covariates are information-rich and potentially high-dimensional. We use the running example of digital advertising in this section, where an advertising platform wants to show an ad from a set of ads to a user. In this example, X_i denotes all the characteristics the advertising platform observes for the user, such as the demographic information and ad response to prior ads. The ad choice is denoted by $w \in \mathcal{W}$ and the outcome of interest can be click or conversion, depending on the platform's objective.

The planner can use a supervised learning algorithm to estimate the outcome by minimizing a loss function that measures how close the actual outcome Y_i is to the estimated outcome $\hat{Y}_i$. Let $\mathcal{L} : \mathcal{Y} \times \hat{\mathcal{Y}} \to \mathbb{R}$ denote the loss function that takes the actual outcome and the estimated outcome as the input and returns a measure of fit. The most common loss functions are the mean squared error when the outcome is continuous and logarithmic loss when the outcome is binary or categorical. Further, let $\mathcal{F}$ denote the class of functions over which we want to learn the function f that estimates the outcome. We can obtain function $\hat{R}$ by optimizing the following objective:

$$\hat{R}(\cdot, \cdot) = \underset{f \in \mathcal{F}}{\operatorname{argmin}} \sum_{i=1}^{N} \mathcal{L}\left(f\left(X_i, W_i\right), Y_i\right) + \lambda \Omega(f), \tag{8}$$

where $f(X_i, W_i) = \hat{Y}_i$ and $\Omega(f)$ is a regularizer that penalizes the variance of the estimator. The planner can use a host of supervised learning algorithms, such as LASSO, Bayesian Additive Regression Trees (BART), XGBoost, and Deep Neural Networks. The choice of which algorithm to use is not *ex-ante* clear. An intuitive solution would be to run all models and choose the one with the best performance on a held-out validation set. However, a key challenge in doing so is that a better out-of-sample loss does not necessarily identify a better personalized policy. We use Equation (3) to illustrate this point. According to Equation (3), we know that $R(X_i, w) = R(X_i, 0) + \tau_w(X_i)$. The first element $R(X_i, 0)$ only depends on X_i, whereas the second element $\tau_w(X_i)$ depends on both the X_i and the treatment. A better fit on the first element does not matter for a better-personalized policy because it would be the same for an individual regardless of the treatment assignment. As a result, using goodness-of-fit measures is not the right practice when researchers want to choose which algorithm to use in this context. See Yoganarasimhan et al. (2022) for an empirical illustration of this issue.

The challenge above raises another important question: is the loss-minimization approach in supervised learning algorithms (Equation (8)) a wrong objective for personalization tasks? The decomposition in Equation (3) helps answer this question: to the extent that the loss-minimization objective helps achieve a calibrated estimate of $\tau_w(\cdot)$, one could rely on it to develop personalized policies. Clearly, the loss-minimization approach in Equation (8) indirectly reduces the loss for $\tau_w(\cdot)$. However, two specific data limitations can lead to miscalibrated estimates of $\tau_w(\cdot)$ as follows:

- Although high-capacity learners can flexibly learn function R, it is important to note that they are designed to accurately predict the outcome for observations drawn from the joint distribution of the training data, i.e., the instance could have been observed in the training data. This requirement highlights a substantial challenge for personalization problems where the planner needs to estimate the outcome for all treatments: if a treatment w could have never been assigned to an observation with covariates X_i, we cannot estimate the outcome for the counterfactual scenario for treatment w in observation with covariates X_i. For example, if an ad for an action game app has never been shown to a 60-year old user, our predictive model cannot estimate the outcome for the scenario where this ad is shown to this user. The decomposition in Equation (3) can help us understand why, as we do not even have an instance of $\tau_w(X_i)$ in $R(X_i, w) = R(X_i, 0) + \tau_w(X_i)$. This makes it impossible for the loss minimization approach to learn τ_w in that instance. Thus, we need to have some randomization in treatment assignments, which is equivalent to the *overlap* assumption in the causal inference literature that requires a non-zero propensity score for all treatments.
- The second challenge relates to how well the learner can distinguish between $R(X_i, 0)$ and $\tau_w(X_i)$. Even if we have some randomization in treatment assignments that creates variation in $\tau_w(X_i)$, it is still unclear if the variation is good enough for the learner to learn $\tau_w(X_i)$ accurately. In particular, if there is a hidden confounding factor that is correlated with both the outcome and treatment assignment, the learner can misattribute the correlation between the confounding factor and outcome to the $\tau_w(X_i)$ rather than $R(X_i, 0)$, resulting in miscalibrated estimates of $\tau_w(X_i)$. An example for this type of misattribution is as follows: suppose that an ad for a gaming app is more likely to be shown to younger users, but age is not one of the variables we observe. Now, if younger users are more likely to click on ads, we may attribute this link to the ad for the gaming app. This is known as the *endogeneity* or *selection on unobservables* problem

in the causal inference literature. If the assignment function $\pi_{\mathcal{D}}$ is known or identifiable, the supervised learning algorithm can better attribute the credit to treatments. This data requirement is equivalent to the *unconfoundedness* in the causal inference literature.

The challenges above highlight that developing good personalized policies is more about data and identification than what algorithm to use, even when using supervised learning algorithms to predict the outcomes. As such, researchers using these algorithms must discuss the data requirements, i.e., whether their setting helps overcome the issues discussed above.

There is a growing literature that applies different outcome modeling approaches to a variety of substantive contexts and examines the returns to personalization – advertising (He et al., 2014; McMahan et al., 2013; Rafieian and Yoganarasimhan, 2021; Rafieian, 2023b), promotions (Simester et al., 2020a,b; Yoganarasimhan et al., 2022), pricing (Kallus and Zhou, 2021; Dubé and Misra, 2021).

3.2 Low-Rank Methods

Suppose the planner has a training data set $\mathcal{D}_{\text{train}} = \{\{(w, Y_i(w))\}_{w \in \mathcal{W}_i}\}_{i=1}^{N}$, where $J = \mathcal{W}$ is very large and there are observed outcomes under multiple treatments in $\mathcal{W}_i$. This immediately presents a very different setting from our general model setup in §2, where the planner can only observe the outcome for one treatment. Further, although side information about X_i can be useful, it is not a necessary data requirement. In this section, we use the running example of movie recommendation to a user, where $w \in \mathcal{W}$ represents movies and $Y_i(w)$ is the rating user i gives to track w, or any other such proxy. In this case, the planner (e.g., Netflix) wants to know whether a user will like an unrated (unseen) movie or not.

Let $Y_{[N \times J]} = [Y_i(j)]$ present the matrix of potential outcomes. The planner has an incomplete version of this matrix $\tilde{Y}_{[N \times J]}$ that is defined as follows:

$$\tilde{Y}_{i,j} = \begin{cases} Y_i\left(w^{(j)}\right) & \text{if } \left(w^{(j)}, Y_i(w^{(j)})\right) \in \mathcal{D}_{\text{train}} \\ ? & \text{otherwise} \end{cases}, \tag{9}$$

where question marks refer to instances where we do not observe the outcome. The planner's personalization problem is to offer a treatment to a given individual i. Outcome modeling in this case translates into completing the observed incomplete matrix $\tilde{Y}_{[N \times J]}$. Intuitively, one could exploit the similarities between individuals or treatments to predict the potential outcomes for the missing entries of matrix $\tilde{Y}_{[N \times J]}$. A common solution to this problem assumes a low-rank structure for the underlying matrix $Y_{[N \times J]}$, which is one of the most common assumptions in the literature on collaborative filtering and recommendation systems (Goldberg et al., 1992; Sarwar et al., 2001;

Linden et al., 2003). Under this assumption, one could use the observed incomplete matrix $\tilde{Y}_{[N \times J]}$ and apply a host of methods to identify two matrices $U_{[R \times N]}$ and $V_{[R \times J]}$ to complete the matrix such that:

$$\hat{Y}_{[N \times J]} = U^T_{[R \times N]} V_{[R \times J]}, \tag{10}$$

where $R \ll \min(N, J)$. The low-rank assumption owes much of its popularity to its simplicity and excellent performance in a variety of application domains. If the underlying individual behavior is a mixture of relatively few factors–as commonly assumed in models of customer segmentation–the underlying potential outcomes matrix will be low-rank. For example, in the context of movie recommendation, there can be different segments that are interested in a few movie styles. Similarly, if treatments are similar enough to be represented as a linear function of a few factors (e.g., movie genres), matrix $Y_{[N \times J]}$ will be low-rank. One of the most canonical use cases of low-rank models is the movie recommendation problem where there are too many users and movies, and the planner (e.g., Netflix) wants to know whether a user will like an unrated (unseen) movie or not. In these cases, the low-rank assumption can be interpreted as follows: there is an R-dimensional representation of latent movie characteristics that explain all the heterogeneity in user preferences. In this setting, one could view matrix $V_{[R \times J]}$ as an R-dimensional representation of all J movies in R features, and matrix $U_{[R \times N]}$ as a representation of all user preferences for these latent features of movies.

Given the importance of the movie recommendation problem for streaming platforms, Netflix launched a million-dollar challenge in 2009-2010, where they asked participants to propose an algorithm that improves the performance of their in-house algorithm by 10% in terms of RMSE. Since then, low-rank methods have received considerable attention in both theoretical and applied domains (Candès and Recht, 2009; Candès and Tao, 2010; Mazumder et al., 2010; Recht, 2011; Koren et al., 2021). Much of the recent work in this domain attempts to find solutions to cases where rich side information about X_i is available, thereby merging the benefits of supervised learning with rich covariate space with that of low-rank methods (Lu et al., 2015; Farias and Li, 2019).

Finally, it is worth noting that, like supervised learning methods in §3.1, low-rank solutions need to adopt a causal view as they can be subject to the same types of selection biases. Schnabel et al. (2016) offer a causal inference view to this problem and propose a debiasing solution based on propensity scoring. Recent work in this domain brings more insights from the causal inference literature and provides solutions to existing challenges (Ma and Chen, 2019; Agarwal et al., 2021; Rafieian, 2023b).

4 Causal Modeling

As discussed in §3, outcome modeling approaches have an important limitation: better prediction of R function does not necessarily produce a better prediction of τ_w, which is what matters for personalized decision-making since it is the treatment-dependent part of the expected reward function. To address this limitation, some approaches focus on developing an unbiased estimation of the the treatment-dependent part of the expected reward function. These approaches can provide theoretical guarantees for the unbiasedness of causal parameters under some assumptions. In §4.1, we present a solution to this problem that uses the assumptions commonly used in the causal inference literature and estimates heterogeneous treatment effects. Next, in §4.2, we discuss the structural modeling solution that has historically been used in marketing and economics.

4.1 Heterogeneous Treatment Effect Estimation

This set of methods views the optimization problem in Equation (5) as follows:

$$\operatorname*{argmax}_{\pi} \mathbb{E}\left[\sum_{w \in \mathcal{W}} \pi(w \mid X_i)\tau_w(X_i)\right], \tag{11}$$

where $\tau_w(X_i)$ is the heterogeneous treatment effect or the Conditional Average Treatment Effect (CATE) at $X = X_i$. As such, the solution to the optimization problem in Equation (11) is to estimate the heterogeneous treatment effects. For any $w \in \mathcal{W}$, our goal is to estimate the conditional average treatment effect $\tau_w(x) = \mathbb{E}[Y(w) - Y(0) \mid X = x]$. There are many solutions proposed in the literature on heterogeneous treatment effect estimation (Rzepakowski and Jaroszewicz, 2012; Athey and Imbens, 2016; Künzel et al., 2017; Shalit et al., 2017; Wager and Athey, 2018; Athey et al., 2019; Farrell et al., 2020; Nie and Wager, 2021; Hitsch et al., 2023). For brevity, we focus on one solution proposed by Nie and Wager (2021) that is based on a general framework that encompasses many features of the prior works on heterogeneous treatment effect estimation.

In this section, we present a general framework to estimate heterogeneous treatment effects. We consider a setting similar to §3.1: the planner has a data set $\mathcal{D} = \{(X_i, W_i, Y_i)\}_{i=1}^N$, where the covariates are information-rich and potentially high-dimensional. For our running example in this section, we use the context of Customer Relationship Management (CRM) campaigns, where firms observe detailed characteristics of individuals such as age and browsing behavior (X_i) and choose from a set of interventions such as free credit and loyalty program (W_i) to achieve higher customer retention (Y_i). We specifically make the following set of assumptions:

Assumption 1. The following three conditions hold for data set $\mathcal{D} = \{(X_i, W_i, Y_i)\}_{i=1}^N$:

1. **Stable Unit Treatment Value Assumption (SUTVA):** The potential outcomes of each individual are not influenced by the treatment assignments or outcomes of other individuals, ensuring independence and non-interference among study units. Further, there are not multiple versions of each treatment that generate different potential outcomes.

2. **Overlap:** The assignment for all individuals is probabilistic. That is, for any treatment $w \in \mathcal{W}$, the propensity score is a number strictly greater than zero and lower than one:

$$0 < \pi_{\mathcal{D}}(w \mid X_i) < 1, \quad \forall w \in \mathcal{W} \tag{12}$$

3. **Unconfoundedness:** The treatment assignment is independent of potential outcomes given observed covariates:

$$W_i \perp\!\!\!\perp \{Y_i(w)\}_{w \in \mathcal{W}} \mid X_i \tag{13}$$

These three assumptions guarantee that the treatment assignment is individualistic, probabilistic, and unconfounded (Imbens and Rubin, 2015). The combination of *overlap* and *unconfoundedness* is what is also known as the *strong ignorability of treatment assignment* (Rosenbaum and Rubin, 1983).

To estimate $\tau_w(\cdot)$ for each $w \in \mathcal{W}$, we can consider a binary treatment case where we only include observations for treatment w and the control treatment. In our running example, one could think of treatment w as free credit and control condition as no intervention. Our data set for treatment w and control will be $\mathcal{D}_w = \{(X_i, W_i, Y_i\}_{W_i \in \{w^{(0)}, w\}}$. Recall Equation (3):

$$Y_i(w) = R(X_i, 0) + \tau_w(X_i) + \epsilon_i(w),$$

We define the dummy variable $D^{(w)} = 1(W = w)$. We can write the observed outcome in our data as follows:

$$Y_i = R(X_i, 0) + D_i^{(w)} \tau_w(X_i) + \epsilon_i, \tag{14}$$

where we use shorthand $\epsilon_i = \epsilon(W_i)$. Since we only focus on the data with two treatments, we re-define the propensity score for treatment w in data $\mathcal{D}_w$ using function e_w, such that $e_w(x) = \Pr(D^{(w)} = 1 \mid X = x, W \in \{w^{(0)}, w\})$. Function $e_w(x)$ is closely related to the propensity score function $\pi_{\mathcal{D}}$ defined in §2 as the difference only comes from the target population.[2]

A fundamental challenge in estimating $\tau_w(x)$ is that the treatment assignment can be a function of x. For example, the firm may choose to give free credit only to customers with a high risk of

[2] It is easy to verify $e_w(x) = \pi_{\mathcal{D}}(w \mid x)/(\pi_{\mathcal{D}}(w \mid x) + \pi_{\mathcal{D}}(w^{(0)} \mid x))$.

churn. This lack of orthogonalization can limit a learner's ability to isolate $\tau_w(X_i)$. We use the unconfoundedness assumption that ensures conditional independence between potential outcomes and treatment given covariates. Under the unconfoundedness assumption, we know that $\mathbb{E}[\epsilon_i \mid X_i, W_i] = \mathbb{E}[\epsilon_i \mid X_i] = 0$. In our example, this assumption implies that conditional on the observables, the assignment to free credit is as good as exogenous. This is a reasonable assumption if we have all the data used for treatment assignment. We define the conditional mean outcome function $m_w(x) = \mathbb{E}[Y \mid X = x, W \in \{w^{(0)}, w\}]$, which is the expected outcome given the information about only the covariates (not treatment assignment). We can use the unconfoundedness assumption to write:

$$m_w(X_i) = \mathbb{E}[R(X_i,0)+D_i^{(w)}\tau_w(X_i)+\epsilon_i(w) \mid X_i = x, W_i \in \{w^{(0)}, w\}] = R(X_i,0)+e_w(X_i)\tau_w(X_i), \tag{15}$$

Now, we can combine Equations (14) and (15) and arrive at the following equation:

$$Y_i - m_w(X_i) = \left(D_i^{(w)} - e_w(X_i)\right)\tau_w(X_i) + \epsilon_i \tag{16}$$

This decomposition–which was first used by Robinson (1988)–has been one of the foundations of much of the recent work at the intersection of machine learning and causal inference. There are multiple reasons for its importance. First, the nuisance functions m_w and e_w capture the dependence of Y and W on X, thereby isolating the treatment effect. This process is what is known as *orthogonalization*, as it orthogonalizes the outcome and treatment with respect to the controls X. Second, both functions can be estimated flexibly with high-capacity machine learning algorithms, which can capture complex selection mechanisms. The better we can learn these functions, the more accurate the treatment effect estimates will be. Most high-capacity learners run the risk of overfitting, so it is crucial for the researchers to perform *cross-fitting*. The idea behind cross-fitting is to split the sample into K folds and estimate functions m_w and e_w for individuals in a specific fold by using the observations in other folds. As such, it is customary to denote the cross-fitted estimates for these functions by $\hat{m}_w^{-k(i)}$ and $\hat{e}_w^{-k(i)}$, where $k : \{1, 2, \ldots, N\} \to \{1, 2, \ldots, K\}$ is a function that determines which fold observation i belongs to and superscript $-k(i)$ refers to the fact that fold $k(i)$ has not been used to estimate the nuisance functions.

Once we have the estimates for the nuisance functions, we can use Equation (16) to form a loss function that minimizes ϵ_i and estimates the function τ from the class of functions $\mathcal{T}$. We can write this loss function as follows:

$$\hat{\tau}_w = \underset{\tau \in \mathcal{T}}{\operatorname{argmin}} \sum_{i \in \mathcal{D}_w} \left(\left(Y_i - \hat{m}_w^{-k(i)}(X_i) \right) - \left(D_i^{(w)} - \hat{e}_w^{-k(i)}(X_i) \right) \tau_w(X_i) \right)^2 + \lambda \Omega(\tau), \tag{17}$$

where Ω regularizes the complexity of the function τ. One could use the same process to obtain $\hat{\tau}_w$ for all $w \in \mathcal{W}$.[3]

Having a loss function for estimating heterogeneous treatment effect estimation is valuable as it allows us to consider a wide class of functions over which we can perform the optimization in Equation 17 and a wide range of problems where we have important constraints to satisfy (e.g., budget, timing).

4.2 Structural Models

In this section, we briefly discuss another solution to the personalization problem that is motivated by causal modeling. In this context, we have some data $\mathcal{D}$ and full knowledge of the underlying data-generating process that is governed by a set of structural parameters. This domain-specific knowledge allows us to make assumptions about the process through which personalization can affect the outcomes. For example, in personalized ranking problems, researchers often make structural assumptions about how exactly consumers search, given a set of structural parameters such as their cost cost (Jeziorski and Segal, 2015). As such, one approach is to assume a structural search model with a set of policy-invariant parameters that determine the user's decision-making process. Identification in these cases is about whether we can identify the structural parameters with the variation at hand. As a result, the data requirements can be even lighter than the ones in Assumption 1. However, it is important to note that the validity of this approach depends heavily on the modeling assumptions made.

5 Direct Policy Learning

The previous section has shown how bringing a causal lens to the problem can help with some challenges in personalization related to confounding and generalizability. However, in many personalization problems, the goal is not to obtain unbiased estimates of the heterogeneous treatment effects, but rather to learn a personalized policy that determines that assignment for each individual. The direct policy learning approach views the optimization problem in Equation (5) as is and attempts to learn a policy that optimizes the target objective. A key advantage of this approach is that we do not need to worry about the unbiasedness of our heterogeneous treatment effect estimates:

[3] Although we focused on the binary case for all treatments, one could use a generalized Robinson decomposition for the case with multiple treatments as in Kaddour et al. (2021).

as long as we find the right policy assignment, it does not matter whether we accurately estimate the treatment effect heterogeneity. Therefore, the direct policy learning approach can give us more flexibility in identifying the optimal personalized policy.

In this section, we present the general idea behind direct policy learning. Recall the optimization problem in Equation (5):

$$\pi^* = \operatorname*{argmax}_{\pi \in \Pi} \rho(\pi)$$

Since we want to find the π that maximizes the true performance function, a natural first step is to think about an unbiased estimate for $\rho(\cdot)$. Importantly, we can use the knowledge of the data-generating policy $\pi_{\mathcal{D}}$ and build an unbiased estimator for the target performance function. This estimator was first proposed by Horvitz and Thompson (1952) and is known as the *Importance Sampling (IS)*, *Inverse Propensity Weighting (IPW)* or *Inverse Propensity Scoring (IPS)* in the literature. We denote the IPS estimator for the performance function by $\hat{\rho}^{IPS}(\cdot)$ and define it as follows:

$$\hat{\rho}^{IPS}(\pi) = \frac{1}{N} \sum_{i=1}^{N} \sum_{w \in \mathcal{W}} \left(\frac{\pi(w \mid X_i)}{\pi_{\mathcal{D}}(w \mid X_i)} 1(W_i = w) Y_i \right) \tag{18}$$

If we use the assumptions used for causal inference as presented in Assumption 1, we can verify the unbiasedness of the IPS estimator as follows:

$$\begin{aligned}
\mathbb{E}[\hat{\rho}^{IPS}(\pi)] &= \mathbb{E}\left[\sum_{w \in \mathcal{W}} \left(\frac{\pi(w \mid X_i)}{\pi_{\mathcal{D}}(w \mid X_i)} 1(W_i = w) Y_i \right) \right] \\
&= \mathbb{E}\left[\sum_{w \in \mathcal{W}} \left(\frac{\pi(w \mid X_i)}{\pi_{\mathcal{D}}(w \mid X_i)} 1(W_i = w) Y_i(w) \right) \right] \\
&= \sum_{w \in \mathcal{W}} \mathbb{E}\left[\left(\frac{\pi(w \mid X_i)}{\pi_{\mathcal{D}}(w \mid X_i)} 1(W_i = w) Y_i(w) \right) \right] \\
&= \sum_{w \in \mathcal{W}} \mathbb{E}\left[\mathbb{E}\left[\frac{1(W_i \mid w)}{\pi_{\mathcal{D}}(w \mid X_i)} \mid X_i \right] \mathbb{E}\left[\pi(w \mid X_i) Y_i(w) \mid X_i \right] \right] \\
&= \sum_{w \in \mathcal{W}} \mathbb{E}\left[\mathbb{E}\left[\pi(w \mid X_i) Y_i(w) \mid X_i \right] \right] \\
&= \sum_{w \in \mathcal{W}} \mathbb{E}\left[\pi(w \mid X_i) Y_i(w) \right] \\
&= \mathbb{E}\left[\sum_{w \in \mathcal{W}} \pi(w \mid X_i) Y_i(w) \right] \\
&= \rho(\pi),
\end{aligned} \tag{19}$$

where the second line uses the SUTVA assumption, and the fourth line uses the unconfoundedness assumption. Since $\hat{\rho}^{IPS}(\cdot)$ is and unbiased estimate of the true performance function, a natural approach to direct policy learning is to find the policy by solving the following optimization problem:

$$\hat{\pi}^* = \underset{\pi \in \Pi}{\operatorname{argmax}} \frac{1}{N} \sum_{i=1}^{N} \sum_{w \in \mathcal{W}} \left(\frac{\pi(w \mid X_i)}{\pi_{\mathcal{D}}(w \mid X_i)} 1(W_i = w) Y_i \right) \tag{20}$$

Many solutions have been proposed to solve this optimization problem in the computer science, statistics, economics, and marketing literature (Manski, 2004; Swaminathan and Joachims, 2015a,b; Kitagawa and Tetenov, 2018; Athey and Wager, 2021; Zhou et al., 2023; Zhang, 2023).

To illustrate the high-level idea behind policy learning and connect it to heterogeneous treatment effects, we present the solution for the case of binary treatments based on Athey and Wager (2021). We want to find π that optimizes $\mathbb{E}[\sum_{w \in \mathcal{W}} \pi(w \mid X_i) Y_i(w)]$. We can define the advantage of policy π relative to the two uniform policies as follows:

$$A(\pi) = 2\rho(\pi) - \mathbb{E}\left[Y_i(0) + Y_i(1)\right] \tag{21}$$

Optimizing $A(\pi)$ is equivalent to the optimizing $\rho(\pi)$ as the second term $\mathbb{E}[Y_i(0) + Y_i(1)]$ is policy-invariant. We can now expand the expression for $A(\pi)$ as follows:

$$\begin{aligned}
A(\pi) &= 2\rho(\pi) - \mathbb{E}\left[Y_i(0) + Y_i(1)\right] \\
&= 2\mathbb{E}\left[\pi(w = 0 \mid X_i) Y_i(0) + \pi(w = 1 \mid X_i) Y_i(1)\right] - \mathbb{E}\left[Y_i(0) + Y_i(1)\right] \\
&= 2\mathbb{E}\left[\left(1 - \pi(w = 1 \mid X_i)\right) Y_i(0) + \pi(w = 1 \mid X_i) Y_i(1)\right] - \mathbb{E}\left[Y_i(0) + Y_i(1)\right] \\
&= \mathbb{E}\left[\left(1 - 2\pi(w = 1 \mid X_i)\right) Y_i(0) + \left(2\pi(w = 1 \mid X_i) - 1\right) Y_i(1)\right] \\
&= \mathbb{E}\left[\left(2\pi(w = 1 \mid X_i) - 1\right)\left(Y_i(1) - Y_i(0)\right)\right] \\
&= \left(2\pi(w = 1 \mid X_i) - 1\right) \mathbb{E}\left[Y_i(1) - Y_i(0) \mid X = X_i\right] \\
&= \left(2\pi(w = 1 \mid X_i) - 1\right) \tau(X_i)
\end{aligned} \tag{22}$$

One could use the IPS estimator and write down $\hat{A}^{IPS}(\pi)$ as follows:

$$\hat{A}^{IPS}(\pi) = \frac{1}{N} \sum_{i=1}^{N} \left(2\pi(w = 1 \mid X_i) - 1\right) \left(\frac{W_i Y_i}{\pi_{\mathcal{D}}(w = 1 \mid X_i)} - \frac{(1 - W_i) Y_i}{1 - \pi_{\mathcal{D}}(w = 1 \mid X_i)} \right) \tag{23}$$

For brevity, let $\Gamma_i = W_i Y_i / \pi_{\mathcal{D}}(w = 1 \mid X_i) - (1 - W_i) Y_i / (1 - \pi_{\mathcal{D}}(w = 1 \mid X_i))$. We can now easily see that Equation (23) is the target function to maximize in a weighted classification problem:

$$\underset{\pi \in \Pi}{\operatorname{argmax}} \hat{A}^{IPS}(\pi) = \underset{\pi \in \Pi}{\operatorname{argmax}} \frac{1}{N} \sum_{i=1}^{N} (2\pi(w = 1 \mid X_i) - 1)\text{sign}(\Gamma_i)|\Gamma_i|, \tag{24}$$

where $\text{sign}(\Gamma_i)$ is the target label and $|\Gamma_i|$ is the target weight. As such, one could use any off-the-shelf classification function to find the policy $\pi(\cdot)$ as a function of x.

6 Evaluation

Once a personalized policy is developed, evaluating its performance is essential before implementation. Evaluation typically involves two main approaches: field experimentation and counterfactual policy evaluation. Field experimentation, considered the gold standard, entails testing the personalized policy against various benchmark policies to estimate total rewards and average treatment effects (ATE), often through A/B testing before implementation on online platforms (Kohavi et al., 2020). However, conducting large-scale field experiments can be costly in terms of time and resources, making alternative evaluation methods necessary. Counterfactual policy evaluation addresses this by assessing what would happen if the personalized policy were adopted, sometimes referred to as off-policy policy evaluation in the literature.

We survey methods for counterfactual policy evaluation that leverage randomization in observed actions and rely on the statistical properties of randomized field experiments to consistently estimate total rewards and ATE under counterfactual scenarios. These methods exploit the idea that, under some degree of randomization, the personalized policy is applied to certain observations in the data purely by chance. Thus, akin to a fully randomized experiment, some observations receive the personalized policy while others do not. However, the assignment may not be entirely exogenous. Nevertheless, if Assumption 1 is satisfied, and propensity scores for observed actions are available, we can reliably estimate the reward for any personalized policy. In the rest of this section, we first present the three standard approaches for off-policy evaluation, following Dudík et al. (2011), and then discuss alternatives used in the literature and the challenges and benefits associated with the different approaches.

6.1 Direct Method

This method estimates the performance of a policy π by directly summing the expected outcomes or rewards of choosing this policy. t Mathematically, it can be expressed as:

$$\hat{\rho}^{\text{DM}}(\pi) = \frac{1}{N} \sum_{i=1}^{N} \sum_{w \in \mathcal{W}} \pi(w \mid X_i) \hat{R}(X_i, w), \tag{25}$$

where $\hat{R}(X_i, w)$ represents the model-based prediction of the outcome for observation i when subjected to the action prescribed by the policy, $\pi(w \mid X_i)$. This technique is termed the Direct Method, as it directly utilizes model predictions in the evaluation process.

The effectiveness of this estimator hinges on the predictive accuracy and counterfactual validity of $\hat{R}$. Generally, it performs well if: (1) the predictive model for $\hat{R}$ exhibits sufficient flexibility (e.g., XGBoost, deep learning methods), and (2) $\hat{R}$ is estimated from a dataset with adequate randomization, meeting the overlap and unconfoundedness assumptions. The latter condition is typically met if propensity scores for all policy-prescribed actions are non-zero and known. However, if an action prescribed by the personalized policy has a zero propensity score (indicating it was never implemented in the data), the predictive model may struggle to accurately estimate its outcome. Thus, scenarios where $\hat{R}$ is learned using flexible machine learning models with inherent randomization in actions are ideal for this estimator. Rafieian and Yoganarasimhan (2021), whose setting satisfies these criteria, conducted an empirical exercise comparing the estimated gains from a model-based approach and found them to closely resemble estimates from a model-free Inverse Propensity Score (IPS) approach (referenced below).

6.2 Inverse Propensity Score Estimator

In contrast to the Direct Method, which relies on a predictive model, the Inverse Propensity Score (IPS) estimator offers a *model-free* approach to evaluation. Originating from the concept of importance sampling by Horvitz and Thompson (1952), it is increasingly utilized in marketing research on personalized policy evaluation. This estimator operates by scaling up observations where users received the policy-prescribed treatment based on their propensity to receive it, creating a pseudo-population that received the treatment. Consequently, the average outcome for this pseudo-population provides an unbiased estimate of the reward for the entire population if the proposed policy were implemented. Formally:

$$\hat{\rho}^{\text{IPS}}(\pi) = \frac{1}{N}\sum_{i=1}^{N}\sum_{w\in\mathcal{W}}\left(\frac{\pi(w \mid X_i)}{\pi_{\mathcal{D}}(w \mid X_i)}1(W_i = w)Y_i\right) \tag{26}$$

where Y_i denotes the observed outcome for observation i, and $\hat{e}(W_i)$ represents the propensity score for action W_i in observation i. In settings where propensity scores are non-zero and known, they can be directly utilized for evaluation. The primary advantage of this approach over the Direct Method is its independence from predictive models for outcome evaluation; instead, it relies on actual outcomes observed in the data. However, its effectiveness depends on the accuracy of propensity scores $\hat{e}(W_i)$. Additionally, even with known but small propensities, the IPS estimator's variance can be high, which can preclude us from making reliable inferences on the relative performance of different personalized policies. Nevertheless, Rafieian and Yoganarasimhan (2021) used both the direct method and IPS approach to counterfactual policy evaluation in a context where both overlap and unconfoundedness assumptions are satisfied and document that both approaches are consistent in the estimated gain from a counterfactual personalized ad targeting policy.

6.3 Doubly Robust Method

The Doubly Robust (DR) estimator builds on the two approaches discussed earlier and overcomes their drawbacks. By leveraging both the Direct Method and the IPS Method, this method ensures that the estimated performance of any policy is correct if either the outcome estimates or propensity scores are accurate. Defined as:

$$\hat{\rho}^{\mathrm{DR}}(\pi) = \frac{1}{N} \sum_{i=1}^{N} \sum_{w \in \mathcal{W}} \left(\frac{\pi(w \mid X_i)}{\pi_{\mathcal{D}}(w \mid X_i)} 1(W_i = w) \left(Y_i - \hat{R}(X_i, w) \right) + \hat{R}(X_i, w) \right) \tag{27}$$

This estimator is particularly advantageous in scenarios where propensities need to be estimated from data, making them potentially noisy, as often seen in observational settings. However, similar to the Direct Method and IPS estimator, if the assumptions of overlap or unconfoundedness are not met, the effectiveness of the DR method is not guaranteed. In such cases, where neither accurate outcome estimation nor accurate propensity score estimation is consistent, the DR method may fail. Recent research aims to enhance these estimators; for instance, Wang et al. (2017) have proposed improvements in this area.

7 Extensions to Other Settings

So far, our discussions were restricted to static settings with a single outcome where the planner can directly assign the treatment and there is no interference between units. Nevertheless, many practical settings do not satisfy these assumptions, and researchers have developed and applied methods that relax one or more of these assumptions. We briefly summarize a few standard ones below without delving into the details and provide references for interested readers.

- *Multi-Objective Personalization:* Most personalization methods help find policies when the planner only cares about a single outcome. However, in reality, most marketing problems deal with multiple objectives. For example, a streaming platform wants good outcomes on both organic and sponsored metrics. The challenge is the conflict between multiple outcomes: optimizing one often comes at the expense of the other outcome. Building on the existing solutions, Rafieian et al. (2023) propose algorithms for the multi-objective personalization problem that recover the Pareto frontier of policies in terms of multiple expected outcomes.

They show that personalization can create value in settings with a strong conflict between multiple objectives and identify policies that substantially improve one outcome without hurting other outcomes.

- *Contextual Bandits:* These classes of models consider settings where the planner does not have access to any existing data and needs to actively learn the model parameters while maximizing profit by playing a good policy (usually regret minimizing). Thus, any (costly) learning today implies the ability to design a better policy in the future, which is referred to as the explore-exploit paradigm Slivkins (2022). In these cases, the planner has access to a set of contextual variables X_i for each unit i and can personalize the policy assignment for each unit to minimize regret. Li et al. (2010) proposed a LinUCB algorithm (short for Linear UCB) for this problem setting and applied it to personalized news recommendations. This framework provides a natural solution to the cold start problem in this setting because new articles, whose valuation and fit we need to learn about, are added regularly, while older articles need less exploration. Researchers have since built on this approach to develop more robust algorithms that do not make the linearity assumption and applied it to several other settings, including advertising and pricing (Aramayo et al., 2023; Jain et al., 2023).
- *Offline Reinforcement Learning:* These classes of models consider scenarios where the planner's objectives are dynamic or forward-looking, i.e., the treatment today affects the states or context variables tomorrow. In these cases, the planner's goal is to choose a policy that maximizes the expected discounted sum of rewards over the horizon. This class of models is specified using the Bellman equation and solved using Q-learning or other value function evaluation approaches (Sutton and Barto, 2018). See Rafieian (2023a) for an application of offline RL to the sequential advertising allocation problem.
- *Settings with compliance issues:* In certain settings, the planner is unable to directly assign a specific treatment to a user. In these settings, the planner can simply take actions that affect treatment intensity but cannot force compliance. In such cases, estimating the returns to personalization has the additional challenge of accounting for not just heterogeneity in treatment effects, but also heterogeneous compliance. Syrgkanis et al. (2019) provides an approach to address this problem using an experiment as an instrument, and Mummalaneni et al. (2022) apply this approach to data from Twitter to recover personalized estimates of users' response to engagement on their content.

8 Conclusion and Directions for Future Research

Many businesses aspire to personalize all aspects of their products and user interactions, including marketing strategies and user experiences. While this goal was once largely aspirational, significant strides in the last decade have made real-time and large-scale personalization increasingly feasible. Two key factors have facilitated these advancements: first, substantial improvements in computing power and data storage, which have enabled the development of powerful machine learning tools capable of personalized, real-time scaling; and second, the concurrent refinement of theoretical and statistical foundations underlying the algorithms and methods used for personalization.

This chapter has surveyed available methods for personalization in static settings, discussed the proper evaluation of personalized policies, and provided examples of personalization across various contexts. However, numerous challenges persist, and achieving effective personalization remains elusive in many practical scenarios. Many exciting avenues of research remain relatively underexplored, e.g., accommodating time drifts, accounting for strategic and/or adversarial agents, and the use of Large Language Models for personalizing content and product recommendations, and we expect these topics to be the focus of interest from both researchers and practitioners in the near future..

References

A. Agarwal, M. Dahleh, D. Shah, and D. Shen. Causal matrix completion. *arXiv preprint arXiv:2109.15154*, 2021.

N. Aramayo, M. Schiappacasse, and M. Goic. A multiarmed bandit approach for house ads recommendations. *Marketing Science*, 42(2):271–292, 2023.

S. Athey and G. Imbens. Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, 113(27):7353–7360, 2016.

S. Athey and S. Wager. Policy learning with observational data. *Econometrica*, 89(1):133–161, 2021.

S. Athey, J. Tibshirani, S. Wager, et al. Generalized random forests. *The Annals of Statistics*, 47(2): 1148–1178, 2019.

E. J. Candès and B. Recht. Exact matrix completion via convex optimization. *Foundations of Computational mathematics*, 9(6):717–772, 2009.

E. J. Candès and T. Tao. The power of convex relaxation: Near-optimal matrix completion. *IEEE Transactions on Information Theory*, 56(5):2053–2080, 2010.

J.-P. Dubé and S. Misra. Personalized pricing and customer welfare. 2021.

M. Dudík, J. Langford, and L. Li. Doubly robust policy evaluation and learning. In *Proceedings* *of the 28th International Conference on International Conference on Machine Learning*, pages 1097–1104. Omnipress, 2011.

D. Dzyabura and H. Yoganarasimhan. Machine learning and marketing. In *Handbook of marketing analytics*, pages 255–279. Edward Elgar Publishing, 2018.

V. F. Farias and A. A. Li. Learning preferences with side information. *Management Science*, 65(7): 3131–3149, 2019.

M. H. Farrell, T. Liang, and S. Misra. Deep learning for individual heterogeneity: An automatic inference framework. *arXiv preprint arXiv:2010.14694*, 2020.

D. Goldberg, D. Nichols, B. M. Oki, and D. Terry. Using collaborative filtering to weave an information tapestry. *Communications of the ACM*, 35(12):61–70, 1992.

X. He, J. Pan, O. Jin, T. Xu, B. Liu, T. Xu, Y. Shi, A. Atallah, R. Herbrich, S. Bowers, et al. Practical Lessons from Predicting Clicks on Ads at Facebook. In *Proceedings of the Eighth International Workshop on Data Mining for Online Advertising*, pages 1–9. ACM, 2014.

G. J. Hitsch, S. Misra, and W. Zhang. Heterogeneous treatment effects and optimal targeting policy evaluation. *Available at SSRN 3111957*, 2023.

D. G. Horvitz and D. J. Thompson. A generalization of sampling without replacement from a finite universe. *Journal of the American statistical Association*, 47(260):663–685, 1952.

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

L. Jain, Z. Li, E. Loghmani, B. Mason, and H. Yoganarasimhan. Effective adaptive exploration of prices and promotions in choice-based demand models. *Available at SSRN 4438537*, 2023.

P. Jeziorski and I. Segal. What makes them click: Empirical analysis of consumer demand for search advertising. *American Economic Journal: Microeconomics*, 7(3):24–53, 2015.

J. Kaddour, Y. Zhu, Q. Liu, M. J. Kusner, and R. Silva. Causal effect inference for structured treatments. *Advances in Neural Information Processing Systems*, 34:24841–24854, 2021.

N. Kallus and A. Zhou. Fairness, welfare, and equity in personalized pricing. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pages 296–314, 2021.

T. Kitagawa and A. Tetenov. Who should be treated? empirical welfare maximization methods for treatment choice. *Econometrica*, 86(2):591–616, 2018.

R. Kohavi, D. Tang, and Y. Xu. *Trustworthy online controlled experiments: A practical guide to a/b testing*. Cambridge University Press, 2020.

Y. Koren, S. Rendle, and R. Bell. Advances in collaborative filtering. *Recommender systems handbook*, pages 91–142, 2021.

S. R. Künzel, J. S. Sekhon, P. J. Bickel, and B. Yu. Meta-learners for estimating heterogeneous treatment effects using machine learning. *arXiv preprint arXiv:1706.03461*, 2017.

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

J. Liaukonyte. Personalized and social commerce. Working Paper, 2021.

G. Linden, B. Smith, and J. York. Amazon. com recommendations: Item-to-item collaborative filtering. *IEEE Internet computing*, 7(1):76–80, 2003.

L. Liu, D. Dzyabura, and N. Mizik. Visual listening in: Extracting brand image portrayed on social media. *Marketing Science*, 39(4):669–686, 2020.

J. Lu, D. Wu, M. Mao, W. Wang, and G. Zhang. Recommender system application developments: a survey. *Decision support systems*, 74:12–32, 2015.

W. Ma and G. H. Chen. Missing not at random in matrix completion: The effectiveness of estimating missingness probabilities under a low nuclear norm assumption. *Advances in neural information processing systems*, 32, 2019.

C. F. Manski. Statistical treatment rules for heterogeneous populations. *Econometrica*, 72(4): 1221–1246, 2004.

R. Mazumder, T. Hastie, and R. Tibshirani. Spectral regularization algorithms for learning large incomplete matrices. *The Journal of Machine Learning Research*, 11:2287–2322, 2010.

H. B. McMahan, G. Holt, D. Sculley, M. Young, D. Ebner, J. Grady, L. Nie, T. Phillips, E. Davydov, D. Golovin, et al. Ad Click Prediction: A View from the Trenches. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1222–1230. ACM, 2013.

S. Mummalaneni, H. Yoganarasimhan, and V. V. Pathak. Producer and consumer engagement on social media platforms. *Available at SSRN 4173537*, 2022.

B. Murthi and S. Sarkar. The role of the management sciences in research on personalization. *Management Science*, 49(10):1344–1362, 2003.

X. Nie and S. Wager. Quasi-oracle estimation of heterogeneous treatment effects. *Biometrika*, 108 (2):299–319, 2021.

D. Proserpio, J. R. Hauser, X. Liu, T. Amano, A. Burnap, T. Guo, D. D. Lee, R. Lewis, K. Misra, E. Schwarz, et al. Soul and machine (learning). *Marketing Letters*, 31(4):393–404, 2020.

O. Rafieian. Optimizing user engagement through adaptive ad sequencing. *Marketing Science*, 42 (5):910–933, 2023a.

O. Rafieian. A matrix completion solution to the problem of ignoring the ignorability assumption. 2023b.

O. Rafieian and H. Yoganarasimhan. Targeting and privacy in mobile advertising. *Marketing Science*, 2021.

O. Rafieian and H. Yoganarasimhan. Ai and personalization. *Artificial Intelligence in Marketing*, pages 77–102, 2023.

O. Rafieian, A. Kapoor, and A. Sharma. Multi-objective personalization of marketing interventions. *Available at SSRN 4394969*, 2023.

B. Recht. A simpler approach to matrix completion. *Journal of Machine Learning Research*, 12 (12), 2011.

P. M. Robinson. Root-n-consistent semiparametric regression. *Econometrica: Journal of the Econometric Society*, pages 931–954, 1988.

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

P. E. Rossi, R. E. McCulloch, and G. M. Allenby. The value of purchase history data in target marketing. *Marketing Science*, 15(4):321–340, 1996.

P. Rzepakowski and S. Jaroszewicz. Decision trees for uplift modeling with single and multiple treatments. *Knowledge and Information Systems*, 32(2):303–327, 2012.

B. Sarwar, G. Karypis, J. Konstan, and J. Riedl. Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th international conference on World Wide Web*, pages 285–295, 2001.

T. Schnabel, A. Swaminathan, A. Singh, N. Chandak, and T. Joachims. Recommendations as treatments: Debiasing learning and evaluation. In *international conference on machine learning*, pages 1670–1679. PMLR, 2016.

U. Shalit, F. D. Johansson, and D. Sontag. Estimating individual treatment effect: generalization bounds and algorithms. In *International Conference on Machine Learning*, pages 3076–3085. PMLR, 2017.

D. Simester, A. Timoshenko, and S. I. Zoumpoulis. Efficiently evaluating targeting policies: Improving on champion vs. challenger experiments. *Management Science*, 66(8):3412–3424, 2020a.

D. Simester, A. Timoshenko, and S. I. Zoumpoulis. Targeting prospective customers: Robustness of machine-learning methods to typical data challenges. *Management Science*, 66(6):2495–2522, 2020b.

A. Slivkins. *Introduction to Multi-Armed Bandits*. 2022.

J. H. Steckel, R. S. Winer, R. E. Bucklin, B. G. Dellaert, X. Drèze, G. Häubl, S. D. Jap, J. D. Little, T. Meyvis, A. L. Montgomery, et al. Choice in interactive environments. *Marketing Letters*, 16

(3):309–320, 2005.

R. S. Sutton and A. G. Barto. *Reinforcement learning: An introduction*. MIT press, 2018.

A. Swaminathan and T. Joachims. Counterfactual risk minimization: Learning from logged bandit feedback. In *International Conference on Machine Learning*, pages 814–823. PMLR, 2015a.

A. Swaminathan and T. Joachims. The self-normalized estimator for counterfactual learning. In *Advances in Neural Information Processing Systems*, pages 3231–3239. Citeseer, 2015b.

V. Syrgkanis, V. Lei, M. Oprescu, M. Hei, K. Battocchi, and G. Lewis. Machine learning estimation of heterogeneous treatment effects with instruments. *Advances in Neural Information Processing Systems*, 32, 2019.

S. Wager and S. Athey. Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 0(0):1–15, 2018. doi: 10.1080/01621459.2017.1319839.

Y.-X. Wang, A. Agarwal, and M. Dudık. Optimal and adaptive off-policy evaluation in contextual bandits. In *International Conference on Machine Learning*, pages 3589–3597. PMLR, 2017.

H. Yoganarasimhan, E. Barzegary, and A. Pani. Design and evaluation of optimal free trials. *Management Science*, 2022.

W. W. Zhang. Optimal comprehensible targeting. 2023.

Z. Zhou, S. Athey, and S. Wager. Offline multi-action policy learning: Generalization and optimization. *Operations Research*, 71(1):148–183, 2023.