ROBUST COUNTERFACTUAL INFERENCES
USING FEATURE LEARNING AND THEIR APPLICATIONS

ABHIMANYU MITRA, KANNAN ACHAN AND SUSHANT KUMAR

Abstract. In a wide variety of applications, including personalization, we want to measure the difference in outcome due to an intervention and thus have to deal with counterfactual inference. The feedback from a customer in any of these situations is only “bandit feedback” - that is, a partial feedback based on whether we chose to intervene or not. Typically randomized experiments are carried out to understand whether an intervention is overall better than no intervention. Here we present a feature learning algorithm to learn from a randomized experiment where the intervention in consideration is most effective and where it is least effective rather than only focusing on the overall impact, thus adding a context to our learning mechanism and extract more information. From the randomized experiment, we learn the feature representations which divide the population into subpopulations where we observe statistically significant difference in average customer feedback between those who were subjected to the intervention and those who were not, with a level of significance \( l \), where \( l \) is a configurable parameter in our model. We use this information to derive the value of the intervention in consideration for each instance in the population. With experiments, we show that using this additional learning, in future interventions, the context for each instance could be leveraged to decide whether to intervene or not.

1. Introduction

One of the most common form of data related to a Web service is customer feedback available in some form of interaction log of customers, when they are interacting with the Web service. However, typically these interaction logs are only partial information, also known as “bandit feedback”, as it is contingent upon the prediction made by the system about what is the best way to present the service or which service to present to the customer; see Swaminathan and Joachims [2015]. For example, in personalization, the prediction is about adapting the content according to the customer, which most often turns out to be picking appropriate contents from a content pool based on customer features like the past browse and purchase activity of the customer. In many of the situations, we want to know how an alternate system of making predictions would have performed which brings us to the realm of counterfactual inference. For example, in personalization, an alternate system could be a different method of picking appropriate contents from the content pool based on customer features.

The problem of counterfactual inference has a rich literature with some of the earlier works dating back to the 1970s and some of the latest appearing in the last few years; see for example Austin [2011], Bang and Robins [2005], Bottou et al. [2013], Chernozhukov et al. [2013], Dudík et al. [2011], Hill [2011], Johansson et al. [2016], Lewis [1973], Rosenbaum and Rubin [1983], Rubin [1974, 2005], Swaminathan and Joachims [2015], van der Laan and Petersen [2007]. However, most of this literature is focussed on observational studies rather than a controlled experiment and dealing with the challenge of eliminating selection bias from the inference. Randomized experiments are known to eliminate this selection bias, but conducting a randomized experiment is costly and it might be even impossible to conduct one such in certain observational studies. Some recent research has been devoted to clever experimental designs which could lead to reducing the cost of the randomized experiment; see Bottou et al. [2013], Johnson et al. [2017], Kohavi et al. [2009], Tang et al. [2010]. In our research, we focus on none of the above problems and accept as our base a randomized experiment framework, where we choose one system as the incumbent and make interventions in that system in randomly chosen situations by overwriting the incumbent system with the predictions of the new system. Thus our base framework already incurs the cost associated with a randomized experiment and therefore for us, a simple comparison of average customer feedbacks (click-through-rates, sales revenue per impression etc.)
for the interventions with those where the incumbent system is not intervened, would yield which system is performing better, free of any selection bias. Therefore, we also do not need sophisticated measurement techniques for counterfactual inference as is required in the observational studies to eliminate selection bias. However, even if we do not have a selection bias, we still might get an inconclusive result due to the noise, when the difference in average customer feedback between those subjected to the intervention and those who were not, is not statistically significantly different from 0, with a level of significance \( l \). Note that the noise is a result of the difference in feedbacks from different instances, mostly due to different preferences as well as different inclinations to provide feedback (some customers are more likely to click than others irrespective of the level of satisfaction with the service). So, we ask the question: what if we group together some of the instances with very similar preferences and very similar inclinations to provide feedback, so that we could get a conclusive result for that group? While a conclusive result for a particular group might not necessarily mean a conclusive global result (it could still be inconclusive at the global level due to noise), we can utilize this information in future system design. In other words, having incurred the cost of a randomized experiment, we ask if we can extract more information from the experiment rather than a global comparison between the two systems? More specifically, we ask the question: can the context of an instance indicate in a robust way (robust against the inherent noise in customer feedback) which system is more suitable rather than picking a system to be used globally (the system which overall performs better)?

We propose a more personalized approach to learn a system’s performance. While one system, say system A, might be overall better than another system, say system B, there might be instances where it is better to predict using system B. The context of an instance might guide us in predicting whether the instance will prefer predictions from system A or predictions from system B. For example, in personalization, we may not yet have found a method of personalizing contents that, based on user features, decides appropriate contents from a content pool, and is universally better than all other methods of personalization. A more realistic scenario is where we have a pool of methods, each of which is best for some considerably large subpopulation, but none of them universally dominates all the others. In such a case, a personalization system which lets the methods divide and conquer, will perform much better than a personalization system which chooses only one of them. A personalization system that lets the individual personalization methods divide and conquer, will work in two layers, where in the first layer of personalization, based on user features (context), the system decides on a method and then, using the chosen method and user features, pick appropriate contents from a content pool for the user. The relative success of this two-layered prediction method when compared to picking the overall better-performing system to be used globally, depends on whether there is enough dissidence about the system preference among the instances. This property of the two-layered prediction is similar to personalization itself, which, compared to a global method of picking best content, works best when the content preferences are vastly different for different instances.

However, finding the set of instances which might prefer a different system than the rest is a combinatorial challenge as the number of subsets explode quickly. Also, unless the subsets are characterized by a function of the contexts of the constituent instances, we could not make the learning useful in future system design. Since the context is usually a feature vector of several dimensions (for example, in user-based personalization, a user’s past browsing and purchase history could become the context), characterizing the subsets of instances where the constituent instances prefer a system that is different than those outside the subset, in terms of the contexts of the constituent instances, is impossible to achieve by iteratively checking each possible function of the contexts of the instances to define each possible subset.

We propose a feature learning algorithm to learn which system makes how much better predictions for what instances compared to the other. As the base of our learning framework, we have a randomized experiment, where there is an incumbent system and interventions are made randomly to overwrite the predictions of the incumbent with the predictions of a new system. We note the “bandit feedback” of the customers for all predictions in the experiment, some with the intervention and some without. For example, the incumbent system could be our current method of personalization, where we choose appropriate content based on user features using the current method, and the intervention is a newly developed method of personalization, which chooses the appropriate contents from the content pool based on user features in a different way.
From the randomized experiment, we learn the feature representations which divide the population of instances into subpopulations where the difference in average customer feedback between those who were subjected to the intervention and those who were not, is statistically significantly different from 0, with a level of significance $l$ and $l$ is a configurable parameter in our model. We use this information to derive the value of the intervention in consideration for each instance in the population based on its context, which we call derived personal valuation, depending on the membership of that particular instance in some subpopulation which exhibited statistically significant valuation, exclusivity of the subpopulation, the estimated average valuation for the subpopulation and its volatility. Note that even though in [Johansson et al. 2016], the authors used feature representations in the problem of counterfactual inference, our motivation for feature learning is completely different from them. In [Johansson et al. 2016], the authors used feature learning to reduce the selection bias in observational studies, whereas in this paper, we start with a randomized experiment which already removes the selection bias and we use feature learning to deduce conclusive results (signal strong compared to noise) for relatively smaller groups, which might be very different from the result at the global level (which could potentially still be inconclusive).

In our above example with two different personalization methods, we infer that the users with positive derived personal valuations (note that the derived personal valuation depends only on the context of the user in the form of user features like past browse and purchase activity) prefer the new method of personalization more than the current one. If the derived personal valuation is negative for everyone, we can safely discard the new method, as the current method universally dominates the new methods. Similarly, if all derived personal valuations are positive, the new method universally dominates the current method and we can safely replace the old method with the new one. However, a more realistic scenario is where the derived personal valuations range from negative to positive values, suggesting that for some users, the current method is better than the new one, whereas for some other users, it is the other way round. In the last and more realistic scenario, we might benefit in keeping the both the methods and build another layer of personalization in the system where based on user features, the system first decides on a method and then using the chosen method and the user features, picks content from the content pool for the user.

Since our method of deriving personal valuation depends only on subpopulations where we have conclusive results (difference in average customer feedback is statistically significantly different from 0, with a level of significance $l$), our derivation of personal valuation at a specified context is more robust. In other words, our derived personal valuation is more free from the inherent noise in customer feedback. The literature on contextual bandit problems is dedicated to building robust estimators for each context (for example, see Dudík et al. [2012], Li et al. [2010, 2011]), but most of them are dedicated to the issue of imbalances in the observed data and the proposed solutions cleverly manage this imbalance. Since we start with a controlled experiment, such imbalances are not primary concern for us. However, our approach to derive robust personal valuations is focussed on the inherent noise in customer feedback and we attempt to make the estimator robust against this inherent noise. Thus our research is fundamentally different from the techniques used in the literature on the contextual bandit problem. To the best of our knowledge, no research has been devoted to the problem we addressed here.

With experiments, we show that the derived personal valuation for each instance could be leveraged in future to decide whether to intervene or not based on the features of the instance.

Figure 1 illustrates an example how the entire process would work. Suppose we have conducted a randomized experiment for 30 days with two versions of a webpage and collected customer clicks on the webpages. We want to understand which version of the webpage generates higher engagement or CTR (click-through-rate). We use, for example, the first 20 days for making context-based robust inferences. This is our set of training instances. We keep the last 10 days to evaluate how a derived personal valuation (DPV) based system design would perform. Note that this is how we will do a system design update based on DPV, where we learn from previous experiments (for example, the training instances) and use the learning to update system design that would impact future predictions (for example, the test instances). Suppose, no conclusion could be inferred at the global level using the training instances, in other words, CTR for version 1 is not statistically significantly different from CTR for version 2 with a level of significance, say 5%. However, using gender of the user as a context gives us more information. Suppose, we find that women, in general, usually
like version 2 better more than version 1 (based on CTR from training instances) and men prefer version 1 over version 2 and both these conclusions could be made with a level of significance, say 5%. If this is our only conclusion, a derived personal valuation (DPV) based system design would suggest showing version 2 to women and version 1 to men. If women in the test instances indeed like version 2 more than version 1, and men in the test instances indeed like version 1 more than version 2, that would validate the fact that a DPV based system would work better than choosing either version1 or version 2 globally. In the second case where we choose one version globally, one of the group (either men or women) would be less engaged. Note that in making the inference that women like version 2 more, we have not used the test instances, but only the training instances. If we indeed used a DPV based system for the test instances, women would only see version 2 and we would not know how engaged they would be with version 1. Thus, for the validation of a DPV based system, it is necessary that the test instances are also part of the randomized experiment, so that we would have women seeing both version 1 and version 2 and would be able to compare the CTR difference without any selection bias and therefore, be able to understand how a DPV based system would perform. After this evaluation of the DPV-based system, if the DPV-based indeed works better, we would update our system using DPV, which in this example is showing version 2 to women and version 1 to men.

2. Mathematical framework

In this section, we provide the criterion for deciding whether a subpopulation exhibits a statistically significant valuation for the intervention in consideration, that is, the difference in average customer feedback is statistically significantly different from 0, for predictions made with the intervention when compared to the ones made without the intervention, with a specified level of significance \( l \) and we configure the parameter \( l \) in our model. Then we construct an optimization problem we need to solve in order to identify subpopulations more likely to pass the criterion. Without loss of generality, we assume each feature could only take a finite number of real values. For features not directly satisfying this assumption, we appropriately merge values or discretize to satisfy this assumption.

To provide a mathematical framework for the problem, let us first introduce some notation. Let \( x_i = (x_{i,1}, x_{i,2}, \cdots, x_{i,F}) \) be the \( i \)-th instance (member of population) and \( x_{i,j} \) be the value of the \( j \)-th feature of the \( i \)-th instance. In other words, \( x_i \) provides the context for instance \( i \). Assume the total number of features used to represent context of an instance is \( F \) and each instance assumes a value for each of the features, that is, a real value of \( x_{i,j} \) is available for each instance \( i \) and each feature \( j \). Let \( y_i \) be the metric or customer feedback for the \( i \)-th instance using which we are measuring valuation of the intervention in consideration, that is, if the intervention impacts positively, the metric is expected to increase and if the
intervention generates negative impact, the metric is expected to decrease. For example, a metric could be the number of clicks per page view. Usually the metric is driven by business goals. We could potentially consider several metrics together making \( y_i \) a vector, but in this paper, we will restrict \( y_i \) to be only a scalar. We assume the metric \( y_i \) for each instance \( i \) are mutually independent (in the probabilistic sense). In future references, we will call the customer feedback we chose to compare the system performances as the metric.

We assume a standard randomized experiment is set-up for the entire population. Thus the population is randomly divided into two groups, a test group \( T \) and a control group \( C \) and while the test group is subjected to predictions with the intervention, the control group is subjected to predictions without the intervention. Also, let us denote the whole population of instances as \( P \). Let \( x_i^T = (x_{i,1}^T, x_{i,2}^T, \ldots, x_{i,F}^T) \) be the feature vector for the \( i \)-th instance of the test group and \( x_j^C = (x_{j,1}^C, x_{j,2}^C, \ldots, x_{j,F}^C) \) be the feature vector of the \( j \)-th instance of the control group. Similarly, let \( y_i^T \) be the metric or customer feedback (clicks, revenue etc.) for the \( i \)-th instance in the test group and \( y_j^C \) be the metric or the customer feedback for the \( j \)-th instance in the control group.

### 2.1. Subpopulation eligibility.

First we establish a criterion for deciding whether a subpopulation shows a statistically significant impact of the intervention in consideration. Since in the randomized experiment set-up, the membership in the test or control group is decided randomly independent of the context \( x \) of the instance, comparison of metrics for test and control groups restricted to any subpopulation based on \( x \), say the subpopulation \( \{ h(x) = v \} \) where \( h(\cdot) : \mathbb{R}^F \rightarrow \mathbb{R}^K \) is a measurable function and \( v \in \mathbb{R}^K \), the measurement of the impact of the intervention in consideration for the subpopulation should also be free of any selection bias. In this paper, we restrict ourselves to linear feature representations and thus for us, \( h(\cdot) = H \times \cdot \), where \( H \) is a \( K \times F \) matrix.

We would like to identify subpopulations where there is a statistically significant impact of the intervention in consideration, or in other words, we want to find subpopulations where the difference in the average metrics for the test and control group restricted to the subpopulation is statistically significantly different from 0, with a level of significance \( l \). Using notation, the condition translates to finding \( H \) and \( v \) such that

\[
\frac{\left| \bar{y}^T H \cdot x = v \right| - \left| \bar{y}^C H \cdot x = v \right|}{\sqrt{\frac{\text{Var}[y^T \mid H \cdot x = v]}{|T \cap \{H \cdot x = v\}|} + \frac{\text{Var}[y^C \mid H \cdot x = v]}{|C \cap \{H \cdot x = v\}|}}}
\times \left( \frac{\left| \bar{y}^T H \cdot x = v \right|^2 - \left| \bar{y}^C H \cdot x = v \right|^2}{\frac{\text{Var}[y^T \mid H \cdot x = v]}{|T \cap \{H \cdot x = v\}|} + \frac{\text{Var}[y^C \mid H \cdot x = v]}{|C \cap \{H \cdot x = v\}|}}\right) > q^2,
\]

(2.1)

where \( \left| \bar{y}^T H \cdot x = v \right| \) and \( \left| \bar{y}^C H \cdot x = v \right| \) are the average metrics from the test and control group respectively restricted to the subpopulation \( \{ H \cdot x = v \} \), \( \text{Var}[y^T \mid H \cdot x = v] \) and \( \text{Var}[y^C \mid H \cdot x = v] \) are the empirical variances of the metric in the test and control group respectively restricted to the subpopulation \( \{ H \cdot x = v \} \), \( q^2 \) is the \((1-l)\)-th quantile of the distribution of the quantity on the LHS of (2.1) under the null hypothesis that the average metric difference between test and control group is 0 (recall, \( l \) is the level of significance of the test, so that under the null hypothesis, the probability of (2.1) being satisfied is exactly \( l \)) and the function \(| \cdot |\) equals the size of the set in its argument. Under the null hypothesis that the intervention has no impact, the distribution of the quantity on the LHS of (2.1) could be approximated by the square of a variable following standard normal distribution.

### 2.2. Finding eligible subpopulations.

We call a subpopulation \( \{ H \cdot x = v \} \) eligible to be included in deriving personal valuation if it satisfies (2.1). Note that satisfying (2.1) is equivalent to rejecting a null hypothesis that the intervention has no impact, where the level of significance of the statistical test is \( l \). We begin by noting that a subpopulation \( \{ H \cdot x = v \} \) may not be eligible according to (2.1) for one of the two reasons: either there is little or no impact of the intervention in consideration, or there is insufficient data to conclude anything, or both. These two reasons, in a way, complement each other. Note that if we have
a lot of data, we could statistically significantly measure even tiny impacts and if we have a small amount of data, the impact needs to be huge in order for us to be able to measure that in a statistically significant way. On the other hand, if we have a huge impact of the intervention in consideration for the subpopulation, all we need is a very small amount of data to measure it in a statistically significant way, and if we have a tiny impact, we need huge amounts of data to do the same. Therefore, the reason for which a subpopulation is not eligible, whether it is the insufficiency of data or the relatively little impact of the intervention in consideration, is more of a subjective decision.

If we could find out a way to identify the subpopulations which has the highest impact without actually checking (2.1), we have satisfied our objective and we need not do anything more. However, no such method is known in its full generality and finding such a method seems to be a harder problem. Instead, we aim to identify subpopulations which have a lot of data, so that when we use (2.1) to measure the impact of intervention in consideration, even relatively smaller impacts could be measured in a statistically significant way.

We would focus on the first term on the LHS of (2.1), which is \( \frac{|T \cap \{H \cdot x = v_h\}| \times |C \cap \{H \cdot x = v_h\}|}{|\{H \cdot x = v_h\}|} \), and is a quantification of the amount of data for the subpopulation \( \{H \cdot x = v_h\} \). This term could be re-written as \( |\{H \cdot x = v_h\}| \cdot w^T \cdot (1 - w^T) \), where \( w^T = \sum_{h \in T \cap \{H \cdot x = v_h\}} 1 \) is the fraction of the subpopulation in the test group. Thus this term is dependent on the subpopulation size \( |\{H \cdot x = v_h\}| \), as well as the fractions of the subpopulation in test and control group, given by \( w^T \) and \( (1 - w^T) \) respectively. We want find \( H \) so as to maximize the first quantity on the LHS of (2.1) for all subpopulations created by \( H \), viz. \( \{H \cdot x = v_h\} \), \( h = 1, 2, \ldots, H \). Since achieving that for all subpopulations created by \( H \) together might not be possible, we want to maximize the expected value of the quantity over all subpopulations created by \( H \). The expected value of the quantity over all subpopulations created by \( H \), where each subpopulation is weighted by its relative size simplifies to

\[
E_{x,h} \left[ \frac{|T \cap \{H \cdot x = v_h\}| \times |C \cap \{H \cdot x = v_h\}|}{|\{H \cdot x = v_h\}|} \right]
\]

\[= \frac{1}{|P|} \sum_{i \in T} \sum_{j \in C} 1(H(x^T_i - x^C_j) = 0),
\]

where \( 1_C = 1 \) if \( C \) is true, and \( 1_C = 0 \) otherwise, are indicator functions indicating whether condition \( C \) is true or not. Recall, the set \( P \) denotes the population of instances. The proof of the equivalence in (2.2) is shown in Appendix A.

So, motivated by (2.2), we search for \( H \) which maximizes the RHS of (2.2). To formulate this as an optimization problem, we define a matrix \( Z \), whose columns are of the form \((x^T_j - x^C_j)\), where \( i \in T \), and \( j \in C \). Thus \( Z \) is a huge matrix with dimensions equal to \( F \times |T| \cdot |C| \). Recall, \( F \) is the total number of features describing instances in the population and \(|T|\) and \(|C|\) are the sizes of the test and control group respectively. Let \( Z_k \) be the \( k \)-th column of \( Z \). Our optimization problem to search for \( H \) is formulated as

\[
\max_{H \{a_k, k = 1, 2, \ldots, |T| \cdot |C|\}} \sum_{k=1}^{T \cdot |C|} a_k \quad \text{s.t.} \quad a_k H \cdot Z_k = 0, \quad a_k \in \{0, 1\}, \quad \forall k = 1, 2, \ldots, |T| \cdot |C|,
\]

Here the variables \( a_k \) act like slack variables, in the sense that if \( H \cdot Z_k \neq 0 \), then \( a_k \) must be 0 in order to satisfy the linear constraint in the optimization problem (2.3). If \( H \cdot Z_k = 0 \), the corresponding slack variable \( a_k \) must assume the value 1 in order to maximize the objective function \( \sum_{k=1}^{T \cdot |C|} a_k \) of the optimization problem (2.3). Therefore, it is easy to see that if \( H, \{a_k, k = 1, 2, \ldots, |T| \cdot |C|\} \) is a solution of the optimization problem (2.3), then \( H \) and \( \{a_k, k = 1, 2, \ldots, |T| \cdot |C|\} \) will satisfy the condition \( \sum_{k=1}^{T \cdot |C|} a_k = \sum_{i \in T} \sum_{j \in C} 1(H(x^T_i - x^C_j) = 0) \), which is equal to \(|P|\) times the RHS of (2.2). So, it follows that \( H \) obtained as a solution from the optimization problem (2.3) would also maximize the LHS of (2.2), which is exactly what we wanted.
Note that here, even if we restrict ourselves to linear feature representations to define the subpopulations, we do not aim to estimate a prediction function for \( y_i \) based on \( x_i \), which is a major focus of methods trying to eliminate selection bias; see Johansson et al. [2016], Rosenbaum and Rubin [1983].

2.3. Reducing the search space of \( H \) matrices. Our next goal is to reduce the search space of \( H \) by eliminating some redundancies and imposing some structure on \( H \) in the optimization problem (2.3). We note that we could demand the rows of \( H \) to be orthonormal without changing the set of subpopulations we consider with the help of the following two propositions. The proofs follow from set equalities and are omitted here for space constraints.

**Proposition 2.1.** The following two statements are true about the set of subpopulations generated by a matrix \( H \):

1. If \( H \) is not full row rank, the set of subpopulations generated by \( H \), viz. \( \{H \cdot x = v_h\}, h = 1, 2, \cdots, H \), could also be generated by a lower dimensional matrix with lesser number of rows.
2. If \( H \) is full row rank, the set of subpopulations generated by \( H \), viz. \( \{H \cdot x = v_h\}, h = 1, 2, \cdots, H \), could also be generated by a matrix with dimensions same as \( H \) and whose rows are orthonormal.

2.4. Searching for multiple \( H \) matrices. Our goal is to find as many subpopulations as possible, which satisfy (2.1), or in other words, find as many subpopulations as possible, where the null hypothesis of no effect of intervention is rejected in the statistical test with level of significance \( l \). The more such subpopulations we find, the more information we extract from the randomized experiment conducted. As discussed previously, the reason for satisfying (2.1) could be attributed to either the amount of data or the magnitude of the impact of the intervention in consideration. In our optimization problem, we focussed on finding subpopulations with the most amount of data. Note that, while it is true that if the amount of data is very little, there is little chance for a subpopulation to satisfy (2.1), but with reasonable amount of data, some subpopulations could still satisfy (2.1) if the magnitude of the impact of the intervention is high enough. It is highly likely that the entire population has the most data, but the impact of the intervention in consideration has no statistically significant impact for the entire population does not preclude the possibility of the impact of the intervention in consideration being statistically significant for a subpopulation with a lot less data.

So, our search for feature representations does not end when we have a solution of optimization problem (2.3) and we keep looking for the next best one, which has less data than the previous one, but still could satisfy (2.1). Here we discuss when we have found a set of matrices (2.3) and we keep looking for the next best one, which has less data than the previous one, but still could satisfy (2.1). Once again, the proof follows from set equalities and is omitted here for space constraints.

**Proposition 2.2.** If \( \mathcal{R}(H_{(n+1)}) \subseteq \mathcal{R}(H_i) \) for some \( i = 1, 2, \cdots, n \), then the set of subpopulations generated by \( H_{(n+1)} \) is the same as the set of subpopulations generated by \( H_i \).

In light of Proposition 2.2, we want to add the restriction that \( \mathcal{R}(H_{(n+1)}) \not\subseteq \mathcal{R}(H_i) \) for \( H_i, i = 1, 2, \cdots, n \). The condition that \( \mathcal{R}(H_{(n+1)}) \not\subseteq \mathcal{R}(H_i) \) could be re-written as the following condition: \( \sum_{j=1}^{K} H_{(n+1),j}^T (I - H_i^T H_i) H_{(n+1),j}^T > 0 \), where \( H_{(n+1),j} \) is the \( j \)-th row of \( H_{(n+1)} \). The equivalence holds since \( H_i^T H_i \) is the projection matrix for \( \mathcal{R}(H_i) \) and hence \( (I - H_i^T H_i) \) is an idempotent matrix. So, putting everything together, having found \( \{H_i, i = 1, 2, \cdots, n\} \), to find the \((n+1)\)-th \( H \) matrix \( H_{(n+1)} \), we solve the following optimization problem:
Note that each run of the optimization problem gives us a feature representation \( \mathbf{H}_{(n+1)} \) and with increasing \( n \), the optimal value of the optimization problem (2.4) drops, indicating that expected quantity of data associated with \( \mathbf{H}_{(n+1)} \) (in the sense of (2.2)) is reducing as \( n \) increases. We stop when \( n \) exceeds a preset threshold or the optimal value drops below a preset threshold. Note that the drop in optimal value of \( \mathbf{H}_{(n+1)} \) with increasing \( n \) is not of concern, because while the drop means there is less expected quantity of data from the subpopulations generated by \( \mathbf{H}_{(n+1)} \) (see (2.2)), ultimately we want to identify all subpopulations which satisfy (2.1) and not only be restricted to subpopulations generated by \( \mathbf{H}_{1} \). The more subpopulations we find that satisfy (2.1), the more information we extract from our randomized experiment.

Note that in earlier discussion, we fixed the dimension of \( \mathbf{H}_{(n+1)} \) as \( K \times F \), where \( F \) is the total number of features defining a context \( \mathbf{x} \). While we cannot change \( F \) as it is given to us, we do have some flexibility in the choice of \( K \). Instead of fixing a particular \( K \), we could start from \( K = 1 \) (dimension of \( \mathbf{H}_{(n+1)} \) is \( K \times F \) ) and then continue increasing \( K \), thus increasing the granularity of the subpopulations. Note that the required magnitude of impact in order to satisfy (2.1) being satisfied for those subpopulations. Thus, it is advisable to keep the \( K \) much lower compared to \( F \). By following this, in our final set of identified subpopulations that satisfy (2.1), some could be generated by \( \mathbf{H}_{(n+1)} \)-s of dimensions \( K_{1} \times F \) and some could be generated by \( \mathbf{H}_{(n+1)} \)-s of dimensions \( K_{2} \times F \), where \( K_{1} \neq K_{2} \). Note that our only aim is to identify as many subpopulations as possible that satisfy (2.1) and we do not care whether they are characterized by \( \mathbf{H} \) matrices of the same dimensions or not.

However, there might be computational limitations as to how long we can prolong our search of \( \mathbf{H} \) matrices and after a while, the expected quantity of data (in the sense of (2.2)) associated with an \( \mathbf{H} \) matrix will become very low. This, in turn, would result in the subpopulations generated by those \( \mathbf{H} \) matrices having lesser and lesser data, which means those subpopulations are more and more unlikely to satisfy (2.1). Thus, we stop when we reach our computational limit or when the expected quantity of data (in the sense of (2.2)) associated with an \( \mathbf{H} \) is low. Even though with this stopping condition, we may have missed some subpopulation which could have satisfied (2.1), in the process, we have extracted a lot more information from the randomized experiment than just the comparison of the average metrics of the test and control group at the global level.

### 3. Algorithm to Find Subpopulations

To solve the optimization problem (2.4), we consider a Lagrangian relaxation of the problem given by

\[
\max_{\mathbf{H}_{(n+1)}, \{a_k \in \{0,1\} : k = 1, 2, \ldots ; |T| \cdot |C|\}} L = \sum_{k=1}^{\max} a_k + \sum_{j=1}^{\frac{|T| \cdot |C|}{|T| \cdot |C|}} \sum_{k=1}^{\frac{|T| \cdot |C|}{|T| \cdot |C|}} \lambda_{j,k} a_k \mathbf{H}_{(n+1),j,k} \cdot \mathbf{Z}_{j,k} +
\]
\[
\sum_{i=1}^{n} \mu_i \left( \sum_{j=1}^{K} H_{(n+1),j}^T \left( I - H_T^j H_j \right) H_{(n+1),j}^T - D \right),
\]

\[s.t. \quad H_{(n+1)} \cdot H_{(n+1)}^T = I,
\]

(3.1)

where \( H_{(n+1),j} \) is the \( j \)-th row of \( H_{(n+1)} \) and \( \mu_i \) and \( \lambda_{j,k} \) are penalty constants; see [Nocedal and Wright 2006]. We take a greedy approach and solve (3.1) by updating \( H_{(n+1)} \) in sequence. We choose \( \mu_i = \frac{1}{2} \) for all \( i = 1, 2, \ldots, n \), and at each update, we change the constant \( \lambda_{j,k} = -\text{sign} \left( H_{(n+1),j}^T Z_k \right) \), where \( H_{(n+1)} \) in the last step is used for computation of the constant.

We update \( H_{(n+1)} \) by gradient descent, where we move the \( H_{(n+1)} \) slightly in the direction of the derivative of \( L \) given in (3.1) w.r.t. \( H_{(n+1)} \). Also, leveraging Proposition 2.2, we could claim that it is good enough to only consider the update vector projected in the orthogonal space of the row space of current \( H_{(n+1)} \). So, finally, the update to the matrix \( H_{(n+1)} \) would be: for a small \( \epsilon > 0 \),

\[
H_{(n+1)}^{updated} = H_{(n+1)} + \epsilon \frac{\partial L}{\partial H_{(n+1)}} \cdot \left( I - H_{(n+1)}^T H_{(n+1)} \right).
\]

(3.2)

The next step is updating \( \{a_k \in \{0, 1\}, k = 1, 2, \ldots, |T| \cdot |C|\} \). To do that, first we compute \( \text{MAX} = \max_{k=1}^{T|C|} \max_{i=1}^{K} \left| H_{(n+1),i}^T Z_k \right| \). Then we update each \( a_k \) in the following way: if the condition \( \max_{i=1}^{K} \left| H_{(n+1),i}^T Z_k \right| > \theta \cdot \text{MAX} \) is satisfied for some value of \( 0 < \theta < 1 \), we set \( a_k = 0 \) and otherwise, we set \( a_k = 1 \). The parameter \( \theta \) could be tuned for the speed of convergence of the algorithm. See Algorithm 1 below for more execution details.

Now we select initializations of the variables. Note that the optimal value of the slack variable \( a_k \) takes the value 1 if and only if the features of the corresponding pair are equal in value once premultiplied by \( x \). We hope that for any initial choice of \( \{a_k \in \{0, 1\}, k = 1, 2, \ldots, |T| \cdot |C|\} \), the appropriate \( H_{(n+1)} \) would be able to do that for every pair and choose \( a_k = 1 \) for \( k = 1, 2, \ldots, |T| \cdot |C| \). For the initial choice of \( H_{(n+1)} \), we will perturb the last found solution \( H_n \) a little as shown below:

\[
H_{(n+1)}^{initial} = H_n + \epsilon \frac{\partial L}{\partial H_n} \cdot \left( I - H_n^T H_n \right).
\]

(3.3)

Note that \( H_n \) is the optimal solution for the optimization problem (2.4) with \( n \) replaced by \( (n-1) \). So, \( H_n \) satisfies all the constraints on \( H_{(n+1)} \) except for the additional constraint imposed when \( n \) is incremented by 1 in the optimization problem (2.4), that is, \( H_{(n+1)} \) cannot belong to \( \mathcal{R}(H_n) \). So, we hope that the perturbation in (3.3) will satisfy all constraints on \( H_{(n+1)} \). For initialization of \( H_1 \), start with a \( K \times K \) identity matrix appended by a zero matrix of dimensions \( K \times (F - K) \).

4. Measuring personal valuation for each instance

The subpopulations we have identified in the previous section might overlap with each other and in this section, we focus on deriving the valuation of the impact for each instance. An instance might be part of several subpopulations which could potentially have different verdicts on the benefits of the intervention.

Some subpopulation that the instance is part of (based on its context), may have a negative effect of the intervention, while some other subpopulation that it is part of, has a positive effect. Thus, given an instance with its context, we need to determine whether intervening with give us better feedback or not. This is what we focus on here.

We assume through the procedures described in the previous sections, we have found a set of subpopulations \( \mathcal{S} \) of the form \( \mathcal{S} = \{H \cdot x = \xi\} \) which satisfy (2.1). We derive the valuation of an instance with context \( x \) as follows:

\[
v(x) = \sum_{\{S:x \in S \text{ and } S \text{ satisfies 2.1}\}} w_{S,x} v(S),
\]

(4.1)
where the weights \(w_{S,x}\) are described below and \(v(S)\) is the average valuation for the subpopulation \(S\) as found from the randomized experiment given by \(v(S) = \bar{y}^T|S| - \bar{y}^C|S|\) where \(\bar{y}^T|S|\) and \(\bar{y}^C|S|\) are the average metrics from the test and control group respectively restricted to the subpopulation \(S\). Also, note that if no subpopulation satisfies the condition in the sum on the RHS of \(4.1\), the value \(v(x)\) is the empty sum, which is 0.

Intuitively, the weights \(w_{S,x}\) should have an inverse relationship with the volatility of the average metric \(v(S)\), as higher volatility means less confidence in our estimate of the average valuation \(v(S)\) for the subpopulation \(S\). Also, the weights \(w_{S,x}\) should penalize bigger subpopulations as they reduce the volatility of \(v(S)\) by adding more data and thus the individual valuations of its members (members of the subpopulation \(S\)) are not necessarily close to the average valuation \(v(S)\) of the subpopulation \(S\). Note that now we are only interested in deriving the valuation of the instance with context \(x\) and not about the average valuation for a subpopulation that it belongs to. Including all these intuitions, we compute the weights by solving the following set of equations:

\[
\sum_{\{S \in \mathcal{S} \text{ and } S \text{ satisfies } (2.1)\}} w_{S,x} = 1, \quad \frac{w_{S,x}}{\sigma[v(S)] \sqrt{|S|}} = \frac{1}{\sqrt{|S|}},
\]

where \(\sigma[v(S)]\) is the volatility of \(v(S)\). Note that the instance represented by its context vector \(x\) plays a role in defining the weights through \(4.1\), where the summands are determined by \(x\).

Note that we could simplify the term that the weights \(w_{S,x}\) given in \(4.2\) are inversely proportional to, as

\[
\sigma[v(S)] \cdot \sqrt{|S|} = \sqrt{\frac{\text{Var}[y^T|S]|T \cap S|}{|T \cap S|} + \frac{\text{Var}[y^C|S]|C \cap S|}{|C \cap S|}} \cdot \sqrt{|S|} = \sqrt{\frac{|C \cap S|}{|S|} \cdot \text{Var}[y^T|S]|T \cap S| + \frac{|T \cap S|}{|S|} \cdot \text{Var}[y^C|S]|C \cap S|} \times \left(\frac{|T \cap S|}{|S|} \cdot \frac{|C \cap S|}{|S|}\right)^{-\frac{1}{2}}.
\]

In a controlled experiment as is our base set-up, we could assume the second term on the RHS of \(4.3\) to be close to a constant for any reasonably large subpopulation and subpopulations need to be reasonably large to be eligible according to \(2.1\). Thus, the weights are dominated by the first term on the RHS of \(4.3\), which means subpopulations where the metrics (customer feedbacks) are less volatile, will get higher weights than those with higher volatility, which conforms with the intuition that we trust those subpopulations more where the metrics are more consistent.
We potentially could derive personal valuation using methods in Johansson et al. [2016], even though it was not the primary objective of the paper. However, note that the authors in Johansson et al. [2016] have used representation learning in removing selection bias, whereas our basic set-up is a randomized experiment and therefore, we do not have any selection bias to begin with. Blindly applying methods of Johansson et al. [2016] on the results of a randomized experiment would result in unnecessary overfitting. Moreover, the methods described in Johansson et al. [2016] would require a known form of prediction function (in Johansson et al. [2016], the authors optimize within a family of prediction functions), whereas we proceed without any assumption on the prediction function and do not even need one for our modeling.

Note that our derivation of personal valuation is based on eligible subpopulations (see (2.1)) we found from the randomized experiment. Thus, in deriving the personal valuation at context $x$, we automatically discard those subpopulations which contain $x$, but where the first order difference (difference in mean) do not rise above the second-order noise (standard deviation), making the derived personal valuation (DPV) at context $x$ more robust. In the literature on contextual bandit problems (for example, see Dudík et al. [2012], Li et al. [2010, 2011]), research has been carried out to reduce bias and variance of estimators for each context $x$, but most of them are dedicated to the issue of imbalance in the training data and the proposed solutions cleverly manage this imbalance. Since our base set-up is a controlled experiment, such imbalances are not primary concern for us. On the other hand, even with the controlled experiment, our conclusions are crippled by the inherent noise in the customer feedbacks, which we attempt to resolve. Thus our approach to derive robust personal valuations is fundamentally different from the techniques used in the literature on contextual bandit problem.

5. Applications

One of the applications is finding relevant target population for a similar future intervention. Given the features for an instance, we can compute derived personal valuations (DPV) even if the instance was not part of the randomized experiment used to identify eligible subpopulations using (2.1). We can assume that a subpopulation with relatively higher DPV will provide more incremental metrics in future interventions than a subpopulation with relatively lower DPV. For example, if the intervention is a new method of personalization as opposed to an existing one, we could show personalized content according to the new method for those who have the positive DPV for this intervention and keep running the current method for the rest. The scope of this application is well beyond personalization, for example, one could use this to target audience for an online ad campaign, where the learning framework is used on a previous similar ad campaign and the intervention is an ad, as opposed to not running any ad campaign at all. In this application, we can target populations with highest DPV to best utilize campaign costs on a receptive audience.

The second application is identifying scope for improvement in the intervention in consideration. The groups with low or negative DPV for a given intervention such as a new method for personalization represent the population for which intervention did not perform well. Thus improvement of the new method of personalization can focus on such groups. Alternatively future interventions can choose to exclude such groups to optimize benefits as suggested before.

5.1. Validation framework. For empirical validation of how a DPV-based system design would work, we propose the following validation framework. We run a randomized experiment where predictions from a system, say, system A, is used as intervention and predictions from another system, say system B, is used as the default option. Customer feedback is collected on all predictions from the experiment. In the validation framework, we divide the randomized experiment data into training and test data instances. We select first 80% of all instances in chronological order as training data and the remaining as test data. We use training data to identify eligible subpopulations which satisfy (2.1) and then use them to compute the DPV for instances in the test data. For each metric, we divide test instances in multiple groups in the order of their DPV by categorizing on quartiles of DPV, everyone below quartile 1 is one group (<Q1), everyone below median but above quartile 1 is another group (Q1- Q2) and so on... Now for each DPV based group in test instances, we note the difference in average metric from those subjected to the intervention and those who were not, and call that our average incremental metric for the group. If DPV derived for the test
instances are actually indicative of how each system will perform compared to the other, we expect to see increasing average incremental metric with increasing DPV. Thus the groups with higher DPV would have more incremental metric than groups with lower DPV. Since the training and test group are separated in a chronological order, this is exactly how we could use DPV in system design, where we learn from our past experiments which system works better for which instances, and use the DPV of future instances to decide the best system for the instance thus optimizing overall performance. Recall that in deriving the DPV for the test instances, we only used the metrics from training instances and the context for test instances, but never used the metrics for the test instances. As shown in Table 1 in our empirical experiment, DPV for the test instances were indeed indicative of system preference of the test instances.

5.2. Results. For our empirical experiment, the intervention was Whole Page Personalization (personalizing different modules of the page together) as opposed to separately personalizing different modules of the webpage. We ran a randomized experiment for 22 days, where a randomly selected fraction of online users were part of the experiment. The users in the randomized experiment were randomly divided into test and control group. The test group was exposed to Whole Page personalization and the control group saw independently personalized modules on the same webpage. We considered click through rate in a particular module category for metric within the web session. The metric assumes value 0 if there was no click for given page view of a module category. We considered several other metrics, where each metric corresponds to clicks restricted to one category of content. We used past site activities of the user in different categories as context/features which are used to characterize an user at the time of the webpage visit (an instance is a user at the time of the webpage visit in this experiment).

We used first 16 days of the randomized experiment as training data and last 6 days as the test data, as suggested in our validation framework. We identified subpopulations satisfying (2.1) from training data and used them to derive personal valuations for users in the test data. Note that the features are determined by user and time-dependent user activity-based features. So, even if the same user comes twice or more during the randomized experiment, they will be treated as different instances. For the experiment below, we fix the level of significance at 30% (recall that the level of significance \( l \) is a configurable parameter in our model).

We present the results for two metrics on the test instances (Table 1): CTR restricted to Category A and Category B modules. For both Category A and B we see that the average CTR difference between test and control groups increased significantly as the DPV increased for the groups. This means that Whole Page Personalization impact increases with increased DPV for the test instances. We only report the results for those DPV-based groups for whom the difference in CTR (restricted to the category) between those receiving Whole Page Personalization and those who did not, was statistically significantly different from 0 with level of significance 30%. Note that due to limitation of data in our test instances, all DPV-based groups in the test instances might not produce conclusive results about system preferences. However, wherever they do, they show DPV being indicative of system preferences.

From the training data (first 16 days of the experiment), the category A CTR (only clicks on content of category A is considered) difference between those who received Whole Page Personalization and those who did not, is -0.13% with standard deviation of 0.23%. So, note that category A CTR is not statistically significantly different for the two groups at a global level. So, in a standard set-up, we would conclude there is no difference in these two experiences in terms of category A CTR metric. However, we found smaller subpopulations with conclusive preferences (restricted to the subpopulations, difference in category A CTR between those who received Whole Page Personalization and those who did not, were statistically significantly different from 0) and using those, we derived the personal valuations. As results in Table 1 illustrate, with category A CTR as the chosen metric, a DPV-based system design would extract much more information from the experiment.

From the training data (first 16 days of the experiment), the category B CTR (only clicks on content of category B is considered) difference between those who received Whole Page Personalization and those who did not, is 0.34% with standard deviation of 0.30%. So, note that category B CTR is statistically significantly different for the two groups at a global level. So, in a standard set-up, we would conclude Whole Page personalization is better in terms of category B CTR metric and impose that on everyone
Table 1. Table for metric - click through rate (CTR)

| Metric     | Group | CTR | Standard deviation of difference in group average CTR | DPV-based groups CTR difference in average CTR | DPV-based group size |
|------------|-------|-----|------------------------------------------------------|-----------------------------------------------|----------------------|
| Category A CTR | Q1    | 0.058 | 0.043 | 160 |
| Category A CTR | Q2-Q3 | 0.135 | 0.108 | 87 |
| Category B CTR | Q1    | -0.032 | 0.030 | 156 |
| Category B CTR | Q1-Q2 | 0.079 | 0.047 | 163 |

(assuming we are only interested in category B CTR). However, as results in Table 1 illustrate, we found a DPV-based group (<Q1) which actually does not prefer Whole Page Personalization (as measured by category B CTR). In this case as well, we would benefit from a DPV based system design.

6. Conclusion

We have proposed a feature learning algorithm to identify optimal system for a given instance based on its context. We have shown that our learning could be leveraged to target populations for future interventions as well as personalize the choice of optimal systems. A framework that leverages such personal preference of optimal systems will generate prediction in a two-layered approach: first choose the preferred system and then make the appropriate prediction using the preferred system. Further research is needed to get clarity on how personalized the choice of systems can be, how many systems a framework can support etc. to build a framework that uses the personalized choice of systems at scale. While we proposed one greedy approach to solve the optimization problem [2,4], further research needs to explore other possibly better ways of solving the optimization problem. Further research could also drive down the computational time. It would also be fruitful to invest in research to estimate noise for DPV and intelligent use of it in system design.

7. Acknowledgement

We sincerely thank Shyam Rapaka, who actively contributed in the material presented in this paper during his tenure at Walmart Labs.

References

Peter C. Austin. 2011. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research* 46, 3 (2011), 399–424.

Heejung Bang and James M. Robins. 2005. Doubly robust estimation in missing data and causal inference models. *Biometrics* 61, 4 (2005), 962–973.

Léon Bottou, Jonas Peters, Joaquin Quiñonero-Candela, Denis X. Charles, Max Chickering, Elon Portugaly, Dipankar Ray, Patrice Y. Simard, and Ed Snelson. 2013. Counterfactual reasoning and learning systems: the example of computational advertising. *The Journal of Machine Learning Research* 14, 1 (2013), 3207–3260.

Victor Chernozhukov, Iván Fernández-Val, and Blaise Melly. 2013. Inference on counterfactual Distributions. *Econometrica* 81, 6 (2013), 2205–2268.

Miroslav Dudík, Dimitru Erhan, John Langford, and Lihong Li. 2012. Sample-efficient nonstationary policy evaluation for contextual bandits. *Proceedings of Uncertainty in Artificial Intelligence (UAI)* (2012), 247–254.

Miroslav Dudík, John Langford, and Lihong Li. 2011. Doubly robust policy evaluation and learning. *Proceedings of the 28th International Conference on International Conference on Machine Learning* (2011), 1097–1104.
Appendix A. Derivation of objective function used in optimization problem formulation

The expected value of the amount of data over all subpopulations created by $H$, where each subpopulation is weighted by its relative size simplifies to

$$E_{v_h} \left[ \frac{|T \cap \{ H \cdot x = v_h \}| \times |C \cap \{ H \cdot x = v_h \}|}{|H \cdot x = v_h|} \right]$$

$$= \sum_{h=1}^{H} \frac{|\{ H \cdot x = v_h \}|}{|P|} \cdot |T \cap \{ H \cdot x = v_h \}| \times |C \cap \{ H \cdot x = v_h \}|$$

$$= \frac{1}{|P|} \sum_{h=1}^{H} \sum_{i \in T} \sum_{j \in C} 1 \{ H \cdot x^i_j = v_h \} \times \left( \sum_{j \in C} 1 \{ H \cdot x^j_h = v_h \} \right),$$

$$= \frac{1}{|P|} \sum_{i \in T} \sum_{j \in C} \sum_{h=1}^{H} 1 \{ H \cdot x^i_j = v_h \} \times 1 \{ H \cdot x^j_h = v_h \};$$

$$= \frac{1}{|P|} \sum_{i \in T} \sum_{j \in C} \sum_{h=1}^{H} 1 \{ H \cdot x^i_j = H \cdot x^j_h \};$$

$$= \frac{1}{|P|} \sum_{i \in T} \sum_{j \in C} 1 \{ H \cdot x^i_j = H \cdot x^j_h \};$$
\( (A.1) \quad = \frac{1}{|\mathcal{P}|} \sum_{i \in \mathcal{P}} \sum_{j \in C} 1\{H(x^T_i - x^C_j) = 0\}, \)

where \(1_C = 1\) if \(C\) is true, and \(1_C = 0\) otherwise, are indicator functions indicating whether condition \(C\) is true or not. Recall, the set \(\mathcal{P}\) denotes the population of instances.

Abhimanyu Mitra, Walmart Labs, 600 West California Avenue, Sunnyvale, CA.
E-mail address: AMitra@walmartlabs.com

Kannan Achan, Walmart Labs, 600 West California Avenue, Sunnyvale, CA.
E-mail address: KAchan@walmartlabs.com

Sushant Kumar, Walmart Labs, 600 West California Avenue, Sunnyvale, CA.
E-mail address: SKumar4@walmartlabs.com