Don’t Throw it Away! The Utility of Unlabeled Data in Fair Decision Making

Miriam Rateike∗
MPI for Intelligent Systems
Tübingen, Germany
Saarland University
Saarbrücken, Germany
mrateike@tue.mpg.de

Ayan Majumdar∗
MPI for Software Systems
Saarland University
Saarbrücken, Germany
ayanm@mpi-sws.org

Olga Mineeva
ETH Zürich
Zürich, Switzerland
MPI for Intelligent Systems
Tübingen, Germany
omineeva@ethz.ch

Krishna P. Gummadi
MPI for Software Systems
Saarbrücken, Germany
gummadi@mpi-sws.org

Isabel Valera
Saarland University
MPI for Software Systems
Saarbrücken, Germany
ivalera@cs.uni-saarland.de

ABSTRACT
Decision making algorithms, in practice, are often trained on data that exhibits a variety of biases. Decision-makers often aim to take decisions based on some ground-truth target that is assumed or expected to be unbiased, i.e., equally distributed across socially salient groups. In many practical settings, the ground-truth cannot be directly observed, and instead, we have to rely on a biased proxy measure of the ground-truth, i.e., biased labels, in the data. In addition, data is often selectively labeled, i.e., even the biased labels are only observed for a small fraction of the data that received a positive decision. To overcome label and selection biases, recent work proposes to learn stochastic, exploring decision policies via i) online training of new policies at each time-step and ii) enforcing fairness as a constraint on performance. However, the existing approach uses only labeled data, disregarding a large amount of unlabeled data, and thereby suffers from high instability and variance in the learned decision policies at different times. In this paper, we propose a novel method based on a variational autoencoder for practical fair decision-making. Our method learns an unbiased data representation leveraging both labeled and unlabeled data and uses the representations to learn a policy in an online process. Using synthetic data, we empirically validate that our method converges to the optimal (fair) policy according to the ground-truth with low variance. In real-world experiments, we further show that our training approach not only offers a more stable learning process but also yields policies with higher fairness as well as utility than previous approaches.

∗Both authors contributed equally to this research.

CCS CONCEPTS
• Computing methodologies → Machine learning algorithms; Online learning settings; • Social and professional topics;

KEYWORDS
fairness, decision making, label bias, selection bias, variational autoencoder, fair representation

1 INTRODUCTION
The extensive literature on fair machine learning has focused primarily on studying the fairness of the predictions by classification models [1, 17, 23, 56, 57] deployed in critical decision-making scenarios. Consider a university admissions process where the goal is to admit students based on their true potential, which may not be directly observable. We assume that independent and identically distributed (i.i.d.) labeled data is available for training. However, in practice, data suffers from selection bias [52] and has been studied extensively in [1, 23, 56, 57]. These fairness studies assume that independent and identically distributed (i.i.d.) labeled data is available for training. However, in decision-making scenarios, data may also suffer from selection bias [52].
bias [36]. That is, we only observe the labels of a small fraction of the data which received positive decisions. For example, we observe (biased) university grades of only the admitted students, resulting in biased, non-i.i.d. labeled data.

In decision-making scenarios affected by both label bias and selection bias, Kilbertus et al. [31] show that to learn the optimal policy, it is necessary to move from learning fair predictions (e.g., predicting grades) to learning fair decisions (e.g., deciding to admit students). To tackle label bias, the authors introduce fairness constraints in the optimization problem. To address selection bias, they propose to learn stochastic, exploring decision policies in an online learning process, where a new decision policy is learned at each time-step. To get unbiased loss estimates from non-i.i.d. labels, the authors further rely on inverse propensity scoring (IPS) [27].

However, the approach ignores unlabeled data. A large fraction of data in the learning process may remain unlabeled due to receiving a negative decision (e.g., students denied admission). Using only labeled data, the approach suffers from high instability and variance in the learning process. In particular, i) the method may give very different outcomes to the same individual, depending on the random initializations of the learning process, and ii) the method may give the same individual very different outcomes at different points in time.

In this paper, we propose a novel online learning process for fair decision-making that leverages both labeled and unlabeled data. Our method learns fair representations of all data using latent variable models in an attempt to capture the unobserved and unbiased ground truth information. In turn, these representations are used to learn a policy that approximates the optimal fair policy (according to the unobserved ground truth). Importantly, as shown in our experiments, our approach leads to a stable, fair learning process, achieving decision policies with similar utility and fairness measures across time and training initializations.

Our primary contributions in this paper are listed below:

1. We propose a novel two-phase decision-making framework that utilizes both labeled and unlabeled data to learn a policy that converges to the optimal (fair) policy with respect to the unobserved and unbiased ground truth.
2. We present a novel policy learning framework, FairAll that relies on a VAE (a latent variable model) architecture to significantly reduce the need for bias correction of selective labeling.
3. Through theoretical analyses and empirical evaluation on synthetic data, we show that the VAE from our FairAll framework is able to learn an unbiased data representation that captures information from the ground truth.
4. Through extensive evaluations based on real-world data, we show that FairAll, compared to prior work, offers a significantly more effective and stable learning process, achieving higher utility and fairness.

1.1 Related Work

Fair Classification. There exists a variety of approaches for fair classification to tackle biased labels. In-processing methods optimize for correct predictions under additional fairness constraints [1, 17, 56, 57]. This requires formulating differentiable fairness constraints and often lead to unstable training [13]. Pre-processing methods instead utilize representation learning first to learn a fair data representation. This representation is then used for downstream predictive tasks [58]. Different methods for fair representation learning have been brought forward, including variational autoencoders (VAEs) [38, 40], normalizing flows [5], and generative adversarial networks [53]. To enforce independence between the learned representation and the sensitive attribute, some methods condition deep generative models on the sensitive features [14, 21, 39, 40], revert to disentanglement [14], perform adversarial training [21, 39, 50] or add regularization, like Maximum-Mean-Discrepancy [21, 38]. While most work on fair representation learning focuses on satisfying group fairness notions [14, 38], some have also considered individual fairness [46] and counterfactual fairness [21]. Recently, contrastive learning for fair representations has attracted much attention [41]. However, it requires the definition of a similarity measure and meaningful data augmentations. This is non-trivial, especially for tabular data. While some recent work [4] exists, further research is needed.

Although all of the works above use fully labeled training data, some have studied fair classification in the presence of partially labeled data [38, 59, 60]. Further, all of these works assume access to a biased proxy label and not the ground truth. Zafar et al. [55] considered analyzing fairness notions separately, assuming access to a ground-truth label. Further, the notions of biased observed proxies and unbiased, unobserved ground-truth (denoted as construct spaces) were discussed in [16, 19]. Note that all of these studies also assume i.i.d. data. But, in most real-world scenarios, the semi-labeled data is not i.i.d. (selection bias [36]). Wick et al. [52] perform an initial fairness-accuracy analysis of classifiers with respect to label and selection bias. Our work, similar to [31] aims to tackle both label and selection bias while transitioning from a static classification setting to online decision-making.

Fair Online Decision Making. Recent works [8, 31] have started exploring fairness in online decision-making processes in the presence of partially labeled non-i.i.d. data. In such settings, convergence to the optimal policy requires exploration and stochastic policies [31]. Kilbertus et al. [31] use an extra fairness constraint in the loss to trade-off between utility and fairness. Additionally, they correct for selection bias in training using inverse propensity scoring (IPS) [27] on the entire loss function, which, unfortunately, can introduce additional variance. Bechavod et al. [8] derive an oracle-efficient bandit algorithm to learn an accurate policy while explicitly controlling the exploration-exploitation trade-off, and thus the variance. However, both approaches disregard a major portion of the data that receives the negative decision and remain unobserved. We show how this data contain useful information about the underlying data distribution. In our approach, we posit utilizing this unlabeled data to reduce the need for IPS. We empirically validate how this helps in faster convergence to an optimal decision policy while providing high utility and fairness during training.\footnote{In Section 5 we compare our method to [31]. Note that [8] is a theoretical work that provides neither experimental results nor an implementation and thus prevents us from comparing to them as the baseline.}
Don’t Throw it Away! The Utility of Unlabeled Data in Fair Decision Making

FAcT ’22, June 21–24, 2022, Seoul, Republic of Korea

2 BACKGROUND AND PROBLEM SETTING

Let us consider a university admission decision-making process inspired by [35], which we will use as a running example. We use uppercase letters for random variables and lowercase letters for their assignments. With $p$, we optionally refer to a probability distribution or a probability mass function. Let $S$ be a random variable indicating a sensitive attribute of an individual describing their membership in a socially salient group (e.g., gender). For simplicity we assume binary $S \in \{-1, 1\}$. Let $X \in \mathbb{R}^n$ be a set of $n$ non-sensitive features that are observed (e.g., high school grades), and may be influenced by $S$. The university aims to take an admission decision $D \in \{0, 1\}$ based on a ground truth target $Y$ (e.g., intellectual potential) [16, 19]. For simplicity we assume $Y \in \{0, 1\}$. Importantly, throughout this paper, we assume that $Y \perp S$ (e.g., potential is equally distributed across social groups), such that an optimal policy decides $D \perp S$.

2.1 Label Bias and Selection Bias

Label Bias. In practice, the ground truth $Y$ often remains unobserved (as it cannot be directly measured). Instead, as shown in Figure 1a, we observe a different label $\tilde{Y}$ (e.g., semester grades) that is assumed to contain information on $Y$ along with measurement noise. We refer to this label as the proxy label. For simplicity, we assume $\tilde{Y} \in \{0, 1\}$. A data-generative process exhibits label bias, if $Y \perp S$, i.e., the proxy target is biased by the sensitive attribute [32]. For example, the same potential may result in lower grades for one demographic group over another due to existing structural discrimination. Figure 1a highlights our assumed data generative process with biased labels. Recall that we aim to take decisions according to $Y \perp S$. However, in the biased label scenario, both $X$ and $\tilde{Y}$ are biased by $S$. A policy that maps $X$ (and potentially $S$) to $\tilde{Y}$ will thus – in the absence of fairness constraints – take biased decisions.

Selection Bias. In practice, algorithms often also need to learn from partially labeled data, where labels $\tilde{Y}$ are observed only for a particular (usually positive) decision. This is called the selective labels problem [36]. For example, a university only knows whether a student gets good semester grades if it accepts the student in the first place. Let these decisions be taken according to policy $\pi$, which may be biased and not optimal. For example, for an individual with features $(x, s)$, a decision $d$ may be taken according to $d \sim \pi(x, s)$.

Then a labeled data point $(x, s, y)$ observed under a policy $\pi$ is not an i.i.d. sample from the true distribution $p(X, S, Y)$. Instead, the data is sampled from the distribution induced by the probability of a positive decision $\pi(d = 1 | x, s)$ such that: $p(x, S, Y) \propto p(y | X, S)\pi(D = 1 | X, S)p(X, S)$ [31].

Corbett-Davies et al. [12] have shown that deterministic decisions (e.g. taken by thresholding) are optimal with respect to $\tilde{Y}$ for i.i.d. data. However, Kilbertus et al. [31] demonstrated that, if labels are not i.i.d., we require exploration, i.e., stochastic decision policies. Such policies map features to a strictly positive distribution over $D$. This implies that the probability of making a positive decision for any individual is never zero. Exploring policies are trained in an online fashion, where the policy is updated at each time step $t$ as $\pi_t$. Moreover, we typically learn a policy from labeled data by minimizing a loss that is a function of the revealed labels. However, if labels are non-i.i.d., it is necessary to perform bias correction to get an unbiased loss estimate. A common technique for such bias correction is inverse propensity score (IPS) weighting [27]. It divides the loss for each labeled datum by the probability with which it was labeled, i.e., received a positive decision under policy $\pi$. However, this bias correction may lead to high variance in the learning process, when this probability is small.

2.2 Measures of Interest: Utility, Fairness, and Their Temporal Stability

Assuming an incurred cost $c$ for every positive decision (e.g., university personnel and facility costs) [12], the decision-maker aims to maximize its profit (revenue – costs), which we call utility. We define utility $U = D(Y < c)$ as a random variable that can take on three values $U \in \{-c, 1-c, 0\}$ depending on decision $D$. A correct positive decision results in a positive profit of $1-c$ (admitting students with high potential leads to more success and funding), an incorrect positive decision results in a negative profit $-c$ (sunk facility costs), and a negative decision (rejecting students) in zero profit. The utility of a policy is then defined as the expected utility $U$ with respect to population $p(X, S, Y)$ and policy $\pi$:

---

**Figure 1:** (a) Ground truth data generative process and (b) our FairAll generative (solid) and inference (dashed) models. Sensitive attribute $S$, non-sensitive attribute $X$, proxy label $\tilde{Y}$, ground truth label $Y$ and decision $D$. Observed (unobserved) random variables grey (white).

\[\text{Note, depending on which label } Y \text{ refers to, } Y \perp S \text{ may not always hold in practice. See Section 6 for a discussion of this assumption.}\]

\[\text{A policy always takes as input features of an individual. Here, we assume the features to be } (X, S). \text{ However, they could also be only } X \text{ or some feature representation.}\]
Definition 2.1 (Utility of a policy [31]). Given utility as a random variable \( U = D(Y - c) \), we define the utility UT of a policy \( \pi \) as the expected overall utility \( UT(\pi) = \mathbb{E}_{X,S,Y} p(X,S,Y) (\pi(D = 1 \mid X, S) (Y - c)) \), where decision and label are \( D, Y \in \{0, 1\} \), and \( c \in (0, 1) \) is a problem specific cost of taking a positive decision.

Note, we defined UT with respect to ground truth target \( Y \). However, as mentioned above, in most practical settings, we only observe proxy \( \tilde{Y} \) and can thus only report UT, i.e., the expected utility measured with respect to proxy \( \tilde{Y} \).

As detailed above, we are interested in taking decisions according to \( Y \), where \( Y \perp S \) (e.g., potential is equally distributed across sensitive groups). A policy that takes decisions based on \( Y \) satisfies counterfactual fairness [35] and demographic parity [17]. This follows directly from the fact that \( Y \) is a non-descendant of \( S \). Any policy \( \pi \) that is a function of the non-descendants of \( S \) (namely \( Y \)) is demographic parity and counterfactually fair [35].\(^4\) The notion of DP fairness for a policy \( \pi \) requires the proportion of decision \( d \) to be the same across all social groups:

\[
\text{Definition 2.2 (Demographic Parity Unfairness of a Policy [17]). We define the demographic parity unfairness (DPU) of a policy \( \pi \) with respect to sensitive attribute \( S \in \{-1, 1\} \) and decision \( D \in \{0, 1\} \) :}
\[
\text{DPU}(\pi) = \mathbb{E}_{X \sim p(X \mid S = 1)} [\pi(D = 1 \mid X, S = 1)] - \mathbb{E}_{X \sim p(X \mid S = -1)} [\pi(D = 1 \mid X, S = -1)]
\]

Correspondingly, a policy is counterfactually fair if it assigns the same decision to an individual in the observed (or, factual) world as well as in a counterfactual world, in which the individual belongs to a different sensitive group.\(^5\)

Definition 2.3 (Counterfactual Unfairness of a Policy [35]). The counterfactual unfairness (CFU) of policy \( \pi \) with respect to a factual individual belonging to \( S = s \) with features \( X^f \) and the decision \( D \in \{0, 1\} \) can be defined as:

\[
\text{CFU}(\pi) = \mathbb{E}_{X^f \sim p(X \mid S = s)} [\pi(D = 1 \mid X^f, s)] - \pi(D = 1 \mid X^{CF}, s')
\]

Here, \( X^f \) refers to the non-sensitive features of an individual in the factual world with sensitive attribute \( s \), and \( X^{CF} \) refers to the non-sensitive features of the same individual in a counterfactual world, where its sensitive attribute is \( s' \) with \( s' \neq s \).

Note, from [35] that satisfying counterfactual fairness implies satisfying demographic parity but not vice-versa. Further, counterfactual analysis requires hypothetical interventions on \( S \) and exact knowledge of the causal generation process. While estimation techniques for real-world data exist [30, 47], in this paper, we only analyze for synthetic data (with access to the true exogenous variables and the structural equations). See Appendix C.4 for more details.

The above metrics allow assessing the performance of one particular policy. However, the online policy learning process outputs, over \( T \) training steps, the set of policies \( \Pi_T^{\pi} = \{\pi_1, \ldots, \pi_T\} \). Assume we wish to stop the learning process from time \( t_1 \). Can we reliably deploy any policy \( \pi_{t_1} \)? Inspired by prior work on temporal fairness [10, 22], we propose a new notion of temporal variance (TV) for a policy learning process. TV indicates how much a metric \( M \) (e.g., utility, fairness) varies for the set of policies across some time interval \([t_1, t_2]\).

\[
\text{Definition 2.4 (Temporal Variance of a Policy Learning Process). We define the temporal variance (TV) of the outcome of a policy learning process } \Pi_t^\pi = \{\pi_t, \ldots, \pi_{t_1}\} \text{ in time interval } [t_1, t_2] \text{ with respect to metric } M \text{ as:}
\]

\[
TV_M(\Pi_t^\pi) = \sqrt{\frac{1}{t_2 - t_1} \sum_{t=t_1}^{t_2} [(M(\pi_t) - \mu_M)^2]}
\]

where \( \mu_M = \frac{1}{t_2 - t_1} \sum_{t=t_1}^{t_2} M(\pi_t) \) denotes the temporal average for the metric \( M \) over the time interval \([t_1, t_2]\).

High TV denotes an unstable learning process, where policies of different time steps achieve different utility and fairness levels for a fixed group of people. For example, policy \( \pi_{t_1} \) may treat the same group of individuals very different compared to \( \pi_{t_1+1} \). A low TV on the other hand indicates a stable learning process, where policies of different time steps achieve similar utility and fairness levels. Hence, it is safe to stop the learning process any time after \( t_1 \). Note, \( \mu_M \) measures the average metric value (e.g. utility, fairness) over the time interval \( t = [t_1, t_2] \).

2.3 Variational Autoencoder

Deep generative models (DGMs), like Variational Autoencoders (VAEs) [33, 45], Normalizing Flows [44] and Generative Adversarial Networks [20] are latent variable models (LVMs) that estimate complex data distributions by capturing hidden structures in the latent space \( Z \).

VAE is one of the most prominent DGMs. It jointly learns a probabilistic generative model \( p_q(X \mid Z) \) (decoder) and an approximate posterior estimator \( q_p(Z \mid X) \) (encoder). Encoder and decoder are parameterized by neural networks. As the marginal likelihood \( p(x) = \int p(x, z) dz \) is intractable, a VAE is trained by maximizing the evidence lower bound (ELBO) of the observations \( x \sim p(X) \), consisting of the expected log-likelihood and the posterior-to-prior KL divergence:

\[
\log p(x) \geq \mathbb{E}_{Z \sim q_p(Z \mid X)} [\log p_q(x \mid z)] - KL(q_p(Z \mid x) || p(Z))
\]

\[
\text{Exp. log-likelihood} \quad \text{KL divergence} \quad = \text{ELBO}(\theta, \phi; x)
\]

3 LEARNING TO DECIDE FAIR

Let us assume that the data \( p(X, S, Y) \) has a generative process as shown in Figure 1a.\(^6\) Recall that a decision-maker ideally aims to take decisions \( d \) according to the unbiased ground truth \( Y \), i.e., \( d \sim p(Y) \) [31]. However, in practice, \( Y \) remains unobserved. Consider

\(^6\)Note, Figure 1a is the same as the causal model presented as the Scenario 3: University success in [35].

\(^1\)Note, this is formulated abusing notation for simplicity. Here, decision \( D \) is a deterministic function of \( Y \), i.e., \( d = \pi(Y) \)).
for now a setting with label bias but no selection bias. We have access to i.i.d. samples from the underlying distribution and the observable label is a biased proxy \( \tilde{Y} \). As per Figure 1a, observed features \( X \) and proxy labels \( \tilde{Y} \) both contain information about \( Y \), but are biased by \( S \) (label bias), i.e., \( X, L, S, \tilde{Y}, LS \). Hence, a policy that takes decisions \( d_{\text{prox}} \sim p(\tilde{Y} | X, S) \) from such biased observed data is unfair.

Assuming access to only biased data, we posit using a conditional latent variable model for fair decision making. As we theoretically show, with the help of a conditional latent variable model (LVM), it is possible to learn a latent representation \( Z \) that: i) is independent of the sensitive \( S \), i.e., \( ZL, S \) and ii) captures the information contained in \( Y \), up to the noise of the observed features and the approximation error of the LVM.

We assume observed features \( X \) and proxy labels \( \tilde{Y} \) are generated by \( Y, S \) and an independent noise variable \( \varepsilon \in \mathbb{R} \).

**Lemma 1.** Assume the observed \( \{X, \tilde{Y}\} \) is a bijective function of the ground-truth \( Y \), sensitive \( S \) and noise \( \varepsilon \), with \( S, Y, \varepsilon \) being pairwise independent. Then, the conditional data entropy is \( H(X, \tilde{Y} | S) = H(Y) + H(\tilde{Y} | S, Y) = H(Y) + H(\varepsilon) \).

So, the conditional data distribution \( p(X, \tilde{Y} | S) \) captures the information of the unobserved ground-truth \( Y \), up to the extent of noise \( \varepsilon \). Next, we consider approximating the underlying data distribution with LVMs.

**Lemma 2.** Given a latent variable model conditional on \( S \) and input data \( \{X, \tilde{Y}\} \), having encoder \( q_{\theta}(Z | X, \tilde{Y}, S) \) and decoder \( p_{\phi}(X, \tilde{Y} | Z, S) \), the mutual information between latent variable \( Z \) and the conditional data distribution \( p(X, \tilde{Y} | S) \) is \( I(Z; X, \tilde{Y} | S) = H(X, \tilde{Y} | S) - \Delta \), where \( \Delta \) is the approximation error of \( Z \).

Hence, the information captured by the latent \( Z \) reduces the uncertainty about the conditional distribution \( p(X, \tilde{Y} | S) \) up to the error. We refer to Appendix A.2 for the detailed proof following [2].

Combining the two lemmas, we get:

\[
I(Z; X, \tilde{Y} | S) = H(Y) + H(\varepsilon) - \Delta
\]

\[\text{information of } \tilde{Y} \text{ noise approx error} \] (1)

The two lemmas together show that using a conditional LVM to model the observed data \( p(X, \tilde{Y} | S) \) allows us to learn a latent variable \( Z \) that captures the information of the unobserved ground-truth \( Y \), up to the extent of noise (note that \( \Delta \) is also dependent on \( \varepsilon \)). Consequently, a policy that learns to make decisions using the latent \( Z \) with respect to the proxy \( \tilde{Y} \) would, in fact, make decisions based on the information contained in \( Y \) (up to the effect of noise \( H(\varepsilon) \)).

As pointed out in Section 2, following [35], a policy deciding based on \( Y \) satisfies counterfactual fairness and demographic parity. Hence, a policy \( \pi \) mapping from \( Z \) to \( \tilde{Y} \) tackles label bias and satisfies both fairness notions without the need for additional constraints (up to the distortion due to \( H(\varepsilon) \) and \( \Delta \)). Following, in Section 4, we propose a pipeline to learn a fair policy using unbiased representations \( Z \) from non-i.i.d. data that suffer from both label and selection bias.

## 4 OUR APPROACH

In this section, we propose a novel online fair policy learning framework FairAll for tackling both biased and selective labels. Our FairAll framework consists of: i) a fair representation learning step that relies on a VAE-based model (illustrated in Figure 1b) trained on both labeled and unlabeled data and: ii) a policy learning approach that leverages the learned fair representations to approximate the optimal fair policy according to the ground truth \( Y \). Both steps of our framework, i.e., the VAE and the policy, are continually optimized as more data becomes available through the development of previous policies (i.e., in an online manner). Note, in taking decisions based on a fair representation, our policy mitigates label bias; in learning a stochastic policy in an online manner, we allow for exploration during training, which mitigates selection bias [31]. Specifically, we correct label bias by conditioning the VAE on sensitive \( S \), while we correct selection bias by weighting our online learning loss with IPS.

In the following, we first detail how to use both labeled and unlabeled data to learn a fair representation and then describe decision learning with policy \( \pi \). Lastly, we present an overview of our entire fair online policy learning pipeline and propose a method to further exploit unlabeled population information.

### 4.1 Learning a Fair Representation

Following the result in Eq. 1, we aim to learn latent \( Z \) that is both informative of \( Y \) and independent of the sensitive attribute \( S \), i.e., \( ZL, S \). Consider an online setting in which we have access to a dataset \( A \) (applicants) at each time step \( t \). The partitioning of \( A \) into labeled data \( A_{UL} \) (accepted applicants) and unlabeled \( A_{UL}^{U} \) (rejected applicants) is invoked by policy \( \pi_t \). To ease notation, we will, in the following, consider a particular time step \( t \) and omit the subscript.

Let us recall that for each data observation, we only observe the proxy label \( \tilde{Y} \) if the previous policy made the positive decision \( D = 1 \) (labeled data), and the actual value of the ground truth remains unobserved. However, we can leverage the fact that the utility with respect to the proxy label \( \tilde{D} = D(\tilde{Y} - c) \), is always observed.\(^8\) This allows us to learn an unbiased latent representation \( Z \) from both labeled and unlabeled data.

**VAE-Based Fair Representation Learning.** Specifically, we build on previous work on semi-supervised and conditional VAEs [32, 44, 49] to approximate the conditional distribution \( p_{\theta}(X, \tilde{U} | S, D) = \int p_{\theta}(X, \tilde{U}, Z | S, D) dZ \) (generative model), and the posterior over the fair latent representation

\[ q_{\omega, \phi}(Z|X,S,D = 1) = \int q_{\omega}(Z|X,S,\tilde{u}, D = 1)q_{\phi}(\tilde{u}|X,S,D = 1)d\tilde{u}. \]

The inference model contains an encoder \( q_{\omega} \) and a separate classifier model \( q_{\phi} \). Note, we condition the inference model on \( D = 1 \) (see Appendix C.1 for an overview) and thus introduce the classifier to predict the label for any unlabeled data point. We optimize the model parameters \((\theta, \phi, \omega)\) by minimizing the following

---

\(^8\)As per Section 2.2, utility \( U \) can take three values dependent on the decision, hence providing a value for accepted and rejected applicants.
objective function:
\[
J(\theta, \phi, \omega) = \alpha \mathbb{E}_{(x, \tilde{u}, s) \sim A^L} [R(\omega; x, s, \tilde{u}, \pi)] + \beta \mathbb{E}_{(x, \tilde{u}, s) \sim A^{UL}, d=1} [L(\theta, \phi; x, s, \tilde{u})] - \mathbb{E}_{(x, s) \sim A^{UL}, d=0} [U(\theta, \phi; x, s)]
\]

IPS-weighted classification loss
\[
- \mathbb{E}_{(x, \tilde{u}, s) \sim A^L, d=1} [L(\theta, \phi; x, s, \tilde{u})] - \mathbb{E}_{(x, s) \sim A^{UL}, d=0} [U(\theta, \phi; x, s)]
\]

Evidence Lower Bound (ELBO)
\[
\log p(x \mid s) = \log p_\theta(x \mid \tilde{u}) - KL(q_\phi(z \mid x, \tilde{u}, d=1) \mid \mid p(z))
\]

Classification Loss.
\[
\mathcal{L}(\theta, \phi; x, s, \tilde{u}) = \mathbb{E}_{z \sim q_\phi(z \mid x, s, \tilde{u}, d=1)}[\log p_\theta(x \mid z, s)] + \log p_\theta(\tilde{u} \mid z, s, D = 1) - KL(q_\phi(z \mid x, s, D = 1) \mid \mid p(z))
\]

For all rejected applicants, which received \(d = 0\), utility is a constant \(U = 0\), such that \(\log p(\tilde{U} = 0)\mid Z, D = 0\) = \(\log 1 = 0\). Thus, the ELBO for unlabeled data is given by:
\[
\mathcal{U}(\theta, \phi; x, s) = \mathbb{E}_{z \sim q_\phi(z \mid x, s, \tilde{u}, d=1)}[\log p_\theta(x \mid z, s)] - KL(q_\phi(z \mid x, s, d = 1 \mid \tilde{u}) \mid \mid p(z))
\]

In our implementation, we learn functions \(p_\theta(X \mid Z, S), p_\theta(\tilde{U} \mid Z, S, D = 1), q_\phi(\tilde{U} \mid X, S, D = 1), q_\phi(Z \mid \tilde{U}, S, X, D = 1)\) with fully-connected (deep) neural networks. See Appendix C and D for practical considerations and training setups.

4.2 Learning a Fair Policy
At each time step after improving our representation learning model, we update the policy \(\pi\) that maps the fair representation \(Z\) into a distribution over decisions \(D\). For each data point in the complete dataset \(A = A^L \cup A^{UL}\) we sample \(z\) using the encoder \(q_\phi(Z \mid \tilde{u}, s, D = 1)\), and use the decoder to get an estimated proxy utility, \(\tilde{u} \sim p_\theta(\tilde{U} \mid Z, S, D = 1)\). Note, for \(D = 1\), the proxy utility is binary. We thus train the policy by minimizing the cross-entropy binary loss: \((1 - \tilde{U}) \log \pi(Z) + \tilde{U} \log (1 - \pi(Z))\). We describe other options for the policy model for our approach in Appendix D.4.

4.3 Exploiting Fully Unlabeled Data
In real-world settings, a decision-maker often has prior access to large unlabeled datasets containing i.i.d. samples from the population of interest. For example, in the case of university admissions, a university may have access to a database of all students who passed their high school diploma to enter university – including those who did not apply at that particular university. Such dataset contains features \(X, S\), but no labels \(Y\), i.e., it is fully unlabeled. However, as we show, such unlabeled data can significantly improve the policy learning process. In particular, the fully unlabeled dataset can help learn fair representations \(Z\) by approximating the conditional \(p(X \mid S)\). That is, we can learn a VAE optimized as:
\[
\mathcal{K}(\theta', \phi'; x, s) = \mathbb{E}_{z \sim q_\phi'(z \mid x, s)}[\log p_\theta(x \mid z, s)] - KL(q_\phi'(z \mid x, s) \mid \mid p(Z))
\]

The resulting unsupervised VAE model can then be used to initialize the parameters of the semisupervised VAE proposed in Section 4.1 via transfer learning. Transfer learning is typically studied in supervised learning [43, 54], where models are pre-trained on large datasets of one domain and then transferred to a different domain with fewer data. In our proposed two-phase approach, we utilize the unlabeled data in a Phase I to initialize our semi-supervised VAE model of the online decision-making Phase II. We initialize parameters \(\phi'\) and \(\theta'\) of semi-supervised VAE with the trained parameters \(\phi'\) and \(\theta'\) of the unsupervised VAE. Note from Eq. 4.1 that the semi-supervised VAE encoder additionally takes \(\tilde{U}\) as input and that the decoder also outputs \(\tilde{U}\). To account for this, we add new neural connections. At the encoder, we add connections from input \(\tilde{U}\) to each neuron in the first hidden layer. At the decoder’s output, we add a new head to output \(\tilde{U}\). The new connections are initialized randomly at the start of Phase II.

4.4 FairAll Overview
Our FairAll learning framework is illustrated in Figure 2 and consists of two phases. In the first step (Phase I), we learn a fair representation using an unsupervised VAE trained in an offline manner using only unlabeled data. We then use the resulting model to initialize the parameters of the semisupervised VAE via transfer learning. In a second step (Phase II), we enter the online decision-learning process, where at each time step, we first update our semisupervised VAE using both labeled and unlabeled data, and then update the decision policy \(\pi\).

5 EXPERIMENTAL RESULTS
In this section, we evaluate our fair policy learning framework with regard to i) its convergence to the optimal policy; ii) its training effectiveness until convergence; and iii) its deployment performance after convergence. We can evaluate (i) on synthetic data only, and evaluate (ii) and (iii) on real-world datasets.
we report (Appendix C). via loss $J_{10}$ dataset. of $f$ effective report $t$ the decision-maker on the training data up to time $t$. ground truth utility $UT$ counterfactual unfairness data with access to the ground truth generative process, we report on both synthetic and real-world datasets. Further, on synthetic and etrier selection (Appendix D.2) and other practical considerations and competing methods (Appendix D.5), details on the hyperparam-

Figure 2: Pipeline of our approach FairAll. In Phase I, we pre-train the VAE using a large pool of unlabeled data with features $x, s$. We transfer the trained parameters $\phi'$ and $\theta'$ of the VAE at the start of Phase II. Next, at each time step $t$, decisions $d$ for a new batch of data are drawn from the current policy $\pi_t$. In case of acceptance ($d = 1$), labels $\hat{y}$ are revealed. The VAE is updated via loss $J$ using both labeled and unlabeled data. Subsequently, the updated VAE outputs fair representations $z$ and estimated utility $\hat{u}$ for all applicants. The policy is updated with these fair representations. In this way, FairAll learns to decide by using all available data.

Baseline and Reference Models. We perform rigorous empirical comparisons among the following learning frameworks:

- **FairAll (I-II)**: Our complete proposed learning framework including Phase I (i.e., offline unsupervised representation learning) and Phase II (online semisupervised representation and policy learning).
- **FairAll (II)**: Baseline approach that make use of only Phase II of the proposed FairAll. This approach allows us to evaluate the impact of Phase I, i.e., fully unlabeled data.
- **FairLab (I-II)**: Baseline approach that consist of unsupervised Phase I and a fully supervised Phase II using only the IPS-weighted ELBO on labeled data. It allows us to evaluate the importance of unlabeled data in Phase II.
- **FairLog [31]**: Competing approach that minimizes the IPS weighted cross entropy (Eq. 5), denoted by $L^{\text{UnfairLog}}$, with a Lagrange fairness constraint, i.e., $L^{\text{UnfairLog}} + \lambda \cdot \text{DPU}$ with DPU as defined in Def. 2.2.
- **UnfairLog [31]**: Unfair reference model, corresponding to FairLog without fairness constraint, i.e., $\lambda = 0$. It allows us to measure the cost of fairness.

We refer to the Appendix for a detailed description of baselines and competing methods (Appendix D.5), details on the hyperparameter selection (Appendix D.2) and other practical considerations (Appendix C).

**Metrics.** We measure observed proxy utility $\hat{UT}$ (Def. 2.1 w.r.t. $\hat{y}$) and demographic parity unfairness DPU (Def. 2.2) on i.i.d. test data on both synthetic and real-world datasets. Further, on synthetic data with access to the ground truth generative process, we report counterfactual unfairness CFU (Def. 2.3)\(^{10}\) as well as the unobserved ground truth utility UT (Def. 2.1 w.r.t. $Y$). For the real-world settings, we report effective $\hat{UT}$ [31], which is the average $\hat{UT}$ accumulated by the decision-maker on the training data up to time $t$. Similarly, we report effective DPU. We also report the temporal variance (Def. 2.4) of $\hat{UT}$ and DPU over time interval $t = [125, 200].$

**Datasets.** We report results on one synthetic and three real-world datasets (more details in Appendix B):

- Synthetic, where $X$ contains 2 features, with Gender $S$ and grades after university admission $Y$ (note that the ground truth intellectual potential $Y$ is considered unobserved and only used for evaluation).
- COMPAS [3, 37], where $X$ contains 3 features, with Race $S$ and no recidivism $Y$.
- CREDIT [15], where $X$ contains 19 non-sensitive features, with Gender $S$ and credit score $Y$.
- MEPS [24], where $X$ contains 39 non-sensitive features, with Race $S$ and high healthcare utilization $Y$.

**Optimal Policies.** In our Synthetic setting, where observed $XLS$ and unobserved $K_S$, the optimal unfair policy (OPT-UNFAIR) takes decisions $d \sim p(Y \mid X, S)$ and the optimal fair policy (OPT-FAIR) decides $d \sim p(Y \mid K)$. OPT-UNFAIR can be approximated with access to i.i.d. samples from the posterior distribution, while OPT-FAIR additionally requires access to unobserved $K$. See Appendix C.3 for details.

**Setup.** We assume access to the proxy labels. Since labeled data is often scarce in practice, we assume an initial HARSH policy which labels around 10-18% of the data that we see in Phase II at $t = 0$. We report results for a lenient policy in Appendix E. For details on initial policies, see Appendix D.1. The decision cost is $c = 0.1$ for MEPS and $c = 0.5$ for all other datasets. See Appendix E.4 for a case study on the impact of the cost value. Wherever applicable, Phase I was trained over a large number of epochs. Phase II was trained for 200 time steps with the same number of candidates in each step. All policy training were done over 10 independent random initializations. For a full description of the experimental setup, see Appendix D. Our code is publicly available\(^{11}\).

---

\(^{10}\)See Appendix C.4 for further details on generating counterfactuals on the synthetic dataset.

\(^{11}\)https://github.com/ayanmaj92/fairall
5.1 Can We Reach the Optimal Fair Policy?

In this section, we use synthetic data to evaluate if, given enough time-steps, the different learning methods yield policies that converge to the optimal (fair) policy (OPT-FAIR) in both terms of proxy utility and fairness.

Results. Figure 3 reports \( \hat{UT} \), DPU and CFU on test data across time. Recall that FairLab (I+II) uses unlabeled data only in Phase I, FairAll (II) only in Phase II, and FairAll (I+II) in both Phase I and Phase II. FairAll (II) starts convergence after approximately 20 steps. FairLab (I+II) starts at higher utility but then exhibits slower convergence behavior and has not fully converged at \( t = 200 \). FairAll (I+II) instead starts from a higher utility and at \( t = 200 \) yields utility and fairness values close to OPT-FAIR. Regarding fairness, FairAll (II) does not converge to OPT-FAIR and exhibits high variance in DPU and CFU both across seeds and over time. Instead, FairAll (I+II) and FairLab (I+II) both rely on unsupervised fair representation learning (Phase I) to converge to a value very close to the optimal one. Thus, the fully unsupervised Phase I appears crucial for convergence to the optimal fair policy with low variance. Comparing FairAll to FairLog, we observe that both methods asymptotically converge to the optimal fair utility. However, FairLog suffers from higher and more noisy DPU while CFU increases over time. This can be explained by the fact that FairLog enforces DPU constraints, and, as shown empirically, DPU does not imply CFU. In contrast, FairAll (I+II) learns a fair representation \( Z \) that achieves CFU and, as a result, also DPU (see Section 3). This evaluation concludes that i) all fair models approximately converge to the optimal utility; ii) unlabeled data helps in faster convergence; iii) utilizing Phase I leads to significantly lower unfairness; iv) our approach FairAll, compared to FairLog, achieves convergence while satisfying both counterfactual and demographic parity fairness notions.

5.2 Do We Actually Trade-off Utility for Fairness?

In Figure 4, we observe that FairAll yields higher fairness but lower observed proxy utility \( \hat{UT} \) compared to the unfair reference model UnfairLog. This observation is often referred to as a fairness-accuracy trade-off [6], which assumes that fairness comes at the cost of utility (or accuracy in predictive settings). As pointed out in [16, 52], if utility is a function of biased labels, then the utility measurement is also biased. Recall the assumption that a decision-maker aims to take decisions based on \( Y \) and aims to maximize UT. For example, potential \((Y = 1)\) drives a successful career, not high university grades \((\hat{Y} = 1)\). In the synthetic setting, we can measure unbiased ground truth utility \( UT \).

Results. In Figure 4, we observe that both FairAll and UnfairLog achieve a similar level of ground truth utility UT, while FairAll reports significantly less discrimination (DPU). This suggests that a decision-maker may not actually trade-off fairness and (true) utility, although we observe lower proxy utility. Note that despite being fair, FairAll (and FairLog) do not reach the UT of the optimal (fair) policy OPT-FAIR. This could be due to the noise \( \epsilon \) in the dataset, which is known to the optimal policy, but is naturally not captured by the models. An in-depth discussion of the phenomenon is outside the scope of this paper, and we refer the reader to the literature [11, 16, 52].

5.3 How Effective Is the Learning Process?

We have shown that FairAll asymptotically outputs a policy that approximates the optimal. Now we investigate how much it costs the decision-maker in terms of utility and the society in terms of fairness to learn this policy. We evaluate the effective proxy UT, and DPU that the online learning process accumulates across time on real-world datasets.

Results. Table 1 summarizes the results for several real-world datasets. FairAll (I+II) consistently accumulates more utility and less unfairness during the online learning process compared to the other approaches. Note that FairLab (I+II), which uses only labeled data in Phase II, accumulates less utility and more unfairness than FairAll (II), which skips Phase I but uses both labeled and unlabeled data in Phase II. This suggests that joint training on labeled and unlabeled data in Phase II significantly improves learning of both a fair representation and policy. Furthermore, FairAll (I+II) outperforms FairAll (II), suggesting that using unlabeled data in
To this end, we compare the performance of the resulting strategy with respect to ground truth unfairness. This suggests that unlabeled data in Phase I benefit fairness, while unlabeled data in Phase II benefit utility. FairAll (I+II) provides significantly higher utility and lower unfairness than the other learning models after approximately 50 time steps. Moreover, FairAll (I+II) even provides higher utility than even the unfair reference model UnfairLog while being as fair as FairLog. This empirically confirms the importance of unlabeled data in both Phase I and Phase II to achieve high test utility and fairness in real-world scenarios.

5.4.1 Can We Reliably Stop Learning at Any Time? Assume that we want to stop the policy learning process from time $t_1$. Can we stop the learning process and deploy the policy at any time $t \geq t_1$? We study this by measuring how much utility (fairness) of the output policies vary over time interval $[t_1, t]$ for a fixed test dataset. A low temporal variance (TV) indicates stable behavior, such that it is safe to terminate the learning process at any time. However, when the variance is high, the decision-maker must carefully select the best stopping point. High TV leads to unstable policies that may lead to low utility and/or high unfairness. In addition, we measure the temporal average $\mu$ of utility (unfairness). It is desirable for a learning process to have both low variance and high (low) $\mu$ for utility (unfairness).

### Table 1: Accumulated utility w.r.t. observed proxy (Effect. $\bar{U}_T$) and demographic parity unfairness (Effect. DPU) measured during the policy learning process for different real-world datasets over time interval $t \in [0, 200]$. We report mean values over the same ten independent seeds. The numbers in the brackets show the standard deviation. All reported values are multiplied by 100 for readability.

<table>
<thead>
<tr>
<th>Model</th>
<th>COMPAS</th>
<th>CREDIT</th>
<th>MEPS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Effect. $\bar{U}_T$</td>
<td>Effect. DPU</td>
<td>Effect. $\bar{U}_T$</td>
</tr>
<tr>
<td>FairAll (I+II)</td>
<td>6.2 (0.8)</td>
<td>10.4 (0.7)</td>
<td>20.3 (0.5)</td>
</tr>
<tr>
<td>FairAll (II)</td>
<td>5.1 (0.6)</td>
<td>10.6 (0.7)</td>
<td>19.8 (1.0)</td>
</tr>
<tr>
<td>FairLab (I+II)</td>
<td>2.3 (0.9)</td>
<td>10.7 (0.9)</td>
<td>16.3 (0.7)</td>
</tr>
<tr>
<td>FairLog</td>
<td>3.6 (0.4)</td>
<td>10.8 (0.9)</td>
<td>19.4 (1.0)</td>
</tr>
<tr>
<td>UnfairLog</td>
<td>4.6 (0.7)</td>
<td>14.6 (1.5)</td>
<td>20.4 (0.5)</td>
</tr>
</tbody>
</table>

Phase I also improves the process. In summary, our results show that unlabeled data available at any learning stage should not be thrown away, but should be used to learn fair and profitable policies. Importantly, our method also outperforms the competing approach FairLog both in terms of utility and fairness. Moreover, even if compared to the unfair approach UnfairLog, FairAll (I+II) accumulates comparable or higher utility with significantly lower unfairness. This suggests that we may not observe a trade-off between observed utility and fairness in the real world, assuming an unbiased ground truth exists. Note that without access to the ground truth label, we cannot comment on the performance with respect to the ground truth utility.

5.4 How Do the Learned Policies Perform During Deployment?

In this section, we analyze how a given strategy, when applied to the population of interest, is expected to perform in the long run. To this end, we compare the performance of the resulting strategy at each time step using an i.i.d. test set.

**Result.** Figure 5 shows how utility $\bar{U}_T$ and unfairness DPU evolve over time for COMPAS. Results on the other real-world datasets are in Appendix E. FairAll (II) achieves both higher utility and higher unfairness compared to FairLab (I+II) with significantly higher variance. This suggests that unlabeled data in Phase I benefit fairness, and that unlabeled data in Phase II benefit utility. FairAll (I+II) provides significantly higher utility and lower unfairness than the other learning models after approximately 50 time steps. Moreover, FairAll (I+II) even provides higher utility than even the unfair reference model UnfairLog while being as fair as FairLog. This empirically confirms the importance of unlabeled data in both Phase I and Phase II to achieve high test utility and fairness in real-world scenarios.
fairness. For CREDIT, for example, note that FairLab (I+II) has a lower TV\textsubscript{DPU} and a higher average unfairness level $\mu\text{DPU}$ than FairAll (I+II). Compared to FairLog, our method FairAll (I+II) is much more stable. It has lower TV\textsubscript{DPU} and TV\textsubscript{UT} as well as a higher average utility $\mu\text{UT}$ and lower unfairness $\mu\text{DPU}$.

6 DISCUSSION OF ASSUMPTIONS AND OUTLOOK
In this section, we discuss the main assumptions, limitations, and potential consequences of our proposed framework.

Assumptions on the Data Generation Process. In this work, we assume that the true data generation process follows Figure 1a and that $S$ is a social construct like gender or race. This follows the understanding that discrimination is based on social constructs and not biological differences [28]. We assume that each individual can be assigned to a social group within a social construct (e.g., gender) and that $S$ is a root node in the graphical model. While being common [35], this is a debated modeling assumption [6, 28]. Furthermore, we assume that we observe a biased proxy variable $\hat{Y}$, and that an unobserved unbiased ground truth $Y$ exists that is independent of $S$. This means that there are no innate differences between social groups with respect to $Y$. However, see Appendix F.1 for examples where $Y \perp S$. We advise any practitioner using our pipeline to carefully assess whether these assumptions hold in their case.

Assumptions of Our Policy Learning Pipeline. First, we assume access to a large unlabeled dataset of i.i.d. samples from the population of interest for our pipeline (Phase I). For example, a university may have access to the grades of all students who graduated from high school in a given time period. We also assume access to sensitive information, which, in the real world, may conflict with privacy regulations (e.g., the principle of data minimization). In this paper, we show that access to a large unlabeled dataset of sensitive information not only increases the utility of the decision-maker but also fairness. We hope that this contributes to the debate between fair algorithmic decision-making and privacy.

Second, we assume a decision-maker has an unlimited budget at each time step, i.e., it can take as many positive decisions as desired. A related line of work [29, 34], deals with fairness in selection problems, where candidates compete for a limited number of positions, or with pipeline settings [18], where candidates enter the decision process one at a time. This is an interesting direction for future research.

Third, we assume that the underlying data distribution does not change over time and thus is not affected by the decisions. This assumption does not necessarily hold in the real world. While it exceeds the scope of this paper, it would be interesting to extend our pipeline to address distribution shift as a consequence of the decision-making process.

Lastly, in Phase II we learn a stochastic policy at each time step and use it to collect new data. We follow Kilbertus et al. [31] in their call for a general discussion about the ethics of non-deterministic decision-making.

Assumptions on Fairness Metrics. While the ethical evaluation of the applicability of a particular fairness notion in a specific context lies outside of our scope, we give an overview of when the use of our pipeline may be helpful in practice. We evaluate fairness based on the demographic parity (DP) notion [17]. Heidari et al. [25] map DP to Rawl’s moral understanding, according to which unequal treatment may only be justified on the basis of innate potential or ambition, not gender or race. Within this framework, the underlying assumption of DP is that individuals should receive utility solely based on the factors for which they are accountable. In this paper, such factors are assumed to be captured by the unobserved ground truth label $Y$. Hertweck et al. [26] show that one should enforce DP not only if socio-demographic groups have similar innate potential at birth, but in some cases even if unjust life biases lead to differences in realized abilities. Wachter et al. [51] similarly argue that DP and counterfactual fairness are bias transforming metrics that acknowledge historical inequalities and assume that certain groups have a worse starting point than others. However, they also warn that, e.g., giving an individual a loan that they cannot repay can exacerbate inequalities.

7 CONCLUSION
In this paper, we considered the problem of learning optimal fair decision policies in the presence of label bias and selection bias.
Table 2: Temporal variance and means of utility ($TV_{\mathbf{UT}}, \mu_{\mathbf{UT}}$) and demographic parity unfairness ($TV_{\mathbf{DPU}}, \mu_{\mathbf{DPU}}$). Metrics are measured on the time interval $t = [125, 200]$ on real-world datasets. We report the mean over ten runs with the standard deviation in brackets. For $TV$, lower values are better. For $\mu$ higher (lower) is better for $UT$ ($DPU$). Values multiplied by 100 for readability.

<table>
<thead>
<tr>
<th></th>
<th>COMPAS</th>
<th></th>
<th>CREDIT</th>
<th></th>
<th>MEPS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$TV_{\mathbf{DPU}}(\mu)$</td>
<td>$\mu_{\mathbf{DPU}}(\mu)$</td>
<td>$TV_{\mathbf{UT}}(\mu)$</td>
<td>$\mu_{\mathbf{UT}}(\mu)$</td>
<td>$TV_{\mathbf{DPU}}(\mu)$</td>
</tr>
<tr>
<td>FairAll (I-II)</td>
<td>1.0 (0.7)</td>
<td>4.3 (3.1)</td>
<td>0.4 (0.3)</td>
<td>8.6 (0.7)</td>
<td>3.0 (2.0)</td>
</tr>
<tr>
<td>FairAll (II)</td>
<td>2.8 (1.7)</td>
<td>7.8 (4.0)</td>
<td>0.8 (0.5)</td>
<td>6.6 (1.0)</td>
<td>2.7 (2.4)</td>
</tr>
<tr>
<td>FairLab (I+II)</td>
<td>0.7 (0.6)</td>
<td>4.3 (3.1)</td>
<td>0.3 (0.3)</td>
<td>4.2 (0.9)</td>
<td>1.7 (1.3)</td>
</tr>
<tr>
<td>FairLog</td>
<td>1.8 (1.4)</td>
<td>4.5 (3.8)</td>
<td>0.5 (0.4)</td>
<td>4.3 (1.1)</td>
<td>3.6 (2.0)</td>
</tr>
</tbody>
</table>

Prior work that attempted to solve the problem by learning stochastic exploratory decision policies in an online process neglects a large amount of unlabeled data and suffers from high variance in the learned policies over time. In this work, we proposed a novel two-phase framework that leverages both labeled and unlabeled data to learn stable, fair decision policies. We introduced a practical sequential decision-making pipeline FairAll that uses the latent representations of a variational autoencoder to learn, over time, the optimal fair policy according to the unobserved ground truth. In line with our assumptions, the decision policies learned by FairAll satisfy both the notion of counterfactual fairness and demographic parity without requiring additional fairness constraints. Through theoretical analysis and experiments with synthetic data, we validate that FairAll converges to the optimal (fair) policy with respect to the unobserved ground truth both in terms of utility and fairness. Compared to the prior work, we show how our modeling approach helps us to be counterfactual and demographic parity fair. On real-world data, we show how FairAll provides not only a significantly more effective learning method, but also higher utility, higher fairness, and a more stable learning process than the existing approach. In comparison to baseline models, we also show the importance of using unlabeled data in both phases to achieve a more accurate, fair, and stable decision learning process.

ACKNOWLEDGMENTS

The authors would like to thank Adrián Javaloy Bornás for his guidance, discussion, and for providing an initial codebase for working on VAEs with heterogenous data. The authors also thank Pablo Sánchez-Martín, Batuhan Koyuncu for providing valuable feedback. Special thanks goes to Diego Baptista Theuerkauf for help with formalizing proofs and notation. Finally, the authors thank the anonymous reviewers for their detailed feedback and comments.

Miriam Rateike is supported by the German Federal Ministry of Education and Research (BMBF): Tübingen AI Center, FKZ: 01IS18039B. Ayan Majumdar and Krishna P. Gummadi are supported by the ERC Advanced Grant “Foundations for Fair Social Computing” (no. 789373). Olga Mineeva is supported by the Max Planck ETH Center for Learning Systems. The authors appreciate formalizing proofs and notation. Finally, the authors thank the anonymous reviewers for their detailed feedback and comments.

REFERENCES


