Learning Cost-Effective and Interpretable Treatment Regimes

Himabindu Lakkaraju
Stanford University

Cynthia Rudin
Duke University

Abstract

Decision makers, such as doctors and judges, make crucial decisions such as recommending treatments to patients, and granting bail to defendants on a daily basis. Such decisions typically involve weighing the potential benefits of taking an action against the costs involved. In this work, we aim to automate this task of learning *cost-effective, interpretable and actionable treatment regimes*. We formulate this as a problem of learning a decision list – a sequence of if-then-else rules – that maps characteristics of subjects (eg., diagnostic test results of patients) to treatments. This yields an end-to-end individualized policy for tests and treatments. We propose a novel objective to construct a decision list which maximizes outcomes for the population, and minimizes overall costs. Since we do not observe the outcomes corresponding to counterfactual scenarios, we use techniques from causal inference literature to infer them. We model the problem of learning the decision list as a Markov Decision Process (MDP) and employ a variant of the Upper Confidence Bound for Trees (UCT) strategy which leverages customized checks for pruning the search space effectively. Experimental results on real world observational data capturing judicial bail decisions and treatment recommendations for asthma patients demonstrate the effectiveness of our approach.

1 Introduction

Medical and judicial decisions can be complex: they involve careful assessment of the subject's condition, analyzing the costs associated with the possible actions, and the nature of the consequent outcomes. Further, there might be costs associated with the assessment of the subject's condition itself (e.g., physical pain endured and monetary costs of medical tests etc.). For instance, a doctor first diagnoses the patient's condition by studying the patient's medical history and ordering a set of relevant tests that are crucial to the diagnosis. In doing so, she also factors in the physical, mental, and monetary costs incurred due to each of these tests. Based on the test results, she carefully deliberates various treatment options, analyzes the potential side-effects as well as the effectiveness of each of these options. Analogously, a judge deciding if a defendant should be granted bail studies the criminal records of the defendant, and enquires for additional information (e.g., defendant's socio-economic status) if needed. She then recommends a course of action that trades off the risk with granting bail to the defendant (the defendant may commit a new crime when out on bail) with the cost of denying bail (adverse effects on defendant or defendant's family, cost of jail to the county).

If Spiro-Test=Pos **and** Prev-Asthma=Yes **and** Cough=High **then** C

Else if Spiro-Test=Pos **and** Prev-Asthma =No **then** Q

Else if Short-Breath =Yes **and** Gender=F **and** Age$\geq$ 40 **and** Prev-Asthma=Yes **then** C

Else if Peak-Flow=Yes **and** Prev-RespIssue=No **and** Wheezing =Yes, **then** Q

Else if Chest-Pain=Yes **and** Prev-RespIssue =Yes **and** Methacholine =Pos **then** C

Else Q

Figure 1: Regime for treatment recommendations for asthma patients output by our framework; Q refers to milder forms of treatment used for quick-relief, and C corresponds to more intense treatments such as controller drugs (C is higher cost than Q); Attributes in blue are least expensive.

In practical situations, human decision makers often leverage personal experience to make decisions, without considering data, even if massive amounts of it exist for the problem at hand. There exist domains where machine learning models could potentially help – but they would need to consider all three aspects discussed above: predictions of counterfactuals, costs of gathering information, and costs of treatments. Further, these models must be interpretable in order to create any reasonable chance of a human decision maker actually using them. In this work, we address the problem of learning such cost-effective, interpretable treatment regimes from observational data.

Proceedings of the 20^{th} International Conference on Artificial Intelligence and Statistics (AISTATS) 2017, Fort Lauderdale, Florida, USA. JMLR: W&CP volume 54.

Prior research addresses various aspects of the problem at hand in isolation. For instance, there exists a large body of literature on estimating treatment effects (namely the causal inference literature), recommending optimal treatments (see, e.g., [1, 33, 7]), and learning intelligible models for prediction (e.g., [18, 15, 21, 4]). However, an effective solution for the problem at hand should ideally incorporate all of the aforementioned aspects. Furthermore, existing solutions for learning treatment regimes neither account for the costs associated with gathering the required information, nor the treatment costs. The goal of this work is to propose a framework which jointly addresses all of the aforementioned aspects.

We address the problem at hand by formulating it as a task of learning a decision list that maps subject characteristics to tests and treatments (such as the one shown in Figure 1) such that it: 1) maximizes the expectation of a pre-specified outcome when used to assign treatments to a population of interest 2) minimizes costs associated with assessing subjects' conditions and 3) minimizes costs associated with the treatments themselves. Note that for each subject in our data, we observe only the outcome for the treatment assigned to that subject, not the counterfactual. We use the doubly robust estimation technique to infer the counterfactuals. We chose decision lists to express the treatment regime because they are interpretable, and allow for tests to be performed sequentially. We propose a novel objective function to learn a decision list optimized with respect to the criteria discussed above. We prove that the proposed objective is NP-hard by reducing it to the weighted exact cover problem. We then optimize this objective by modeling it as a Markov Decision Process (MDP) and employing a variant of the Upper Confidence Bound for Trees (UCT) strategy which leverages customized checks for pruning the search space effectively.

We empirically evaluate the proposed framework on two real world datasets: 1) judicial bail decisions 2) treatment recommendations for asthma patients. Our results demonstrate that the regimes output by our framework result in improved outcomes compared to state-of-the-art baselines at much lower costs. The treatment regimes we found are not complicated and require few diagnostic checks to determine the optimal treatment.

2 Related Work

Below, we provide an overview of related research on learning treatment regimes, subgroup analysis, and interpretable models.

Treatment Regimes. The problem of learning treatment regimes has been extensively studied in the context of medicine and health care. Along the lines of [39], literature on treatment regimes can be categorized as: *regression-based methods* and *policy-search-based methods*. *Regression-based methods* [27, 31, 28, 32, 43, 27, 25] model the conditional distribution of the outcomes given the treatment and characteristics of patients and choose the treatment resulting in the best possible outcome for each individual. *Policy-search-based methods* [28, 42, 41, 38, 39] search for a policy (a function which assigns treatments to individuals) within a pre-specified class of policies such that the resulting expected outcome is maximized across the population of interest. Furthermore, recent research in personalized medicine has also focused on developing *dynamic treatment regimes* [14, 40, 33, 7] where the goal is to learn treatment regimes that maximize outcomes for patients in a given population by recommending a sequence of appropriate treatments over time, based on the state of the patient. Very few of the aforementioned solutions [39, 24, 40] produce regimes that are intelligible. None of the aforementioned approaches explicitly account for treatment costs and costs associated with gathering information pertaining to patient characteristics.

While most work on learning treatment regimes has been done in the context of medicine, the same ideas apply to policies in other fields. To the best of our knowledge, this work is the first attempt in extending work on treatment regimes to judicial bail decisions.

Subgroup Analysis. The goal of this line of research is to find out whether there exist subgroups of individuals in which a given treatment exhibits heterogeneous effects, and if so, how the treatment effect varies across them. This problem has been well studied [30, 8, 19, 3, 9, 34]. However, identifying subgroups with heterogeneous treatment effects does not readily provide us with regimes.

Interpretable Models. A large body of machine learning literature focuses on developing interpretable models for classification [18, 35, 15, 21, 4] and clustering [11, 17, 16]. To this end, various classes of models such as decision lists [18], decision sets [15, 36], prototype (case) based models [4], and generalized additive models [21] were proposed. These classes of models were not conceived to model treatment effects. There has been recent work on leveraging decision lists to describe estimated treatment regimes [34, 24, 13, 39]. These solutions do not account for the treatment costs or costs involved in gathering patient characteristics. Several of these techniques use greedy methods, which causes issues with the quality of the models. Lastly, there has also been some work on cost-sensitive learning of classification models such as decision trees [20, 10]. This is, however, not applicable to the problem at hand because it does not model treatment effects.

3 Our Framework

First, we formalize the notion of treatment regimes and discuss how to represent them as decision lists. We then propose an objective function for constructing cost-effective and interpretable treatment regimes.

3.1 Input Data and Cost Functions

Consider a dataset $\mathcal{D} = \{(\mathbf{x}_1, a_1, y_1), (\mathbf{x}_2, a_2, y_2) \cdots (\mathbf{x}_N, a_N, y_N)\}$ comprised of N independent and identically distributed observations, each of which corresponds to a *subject* (individual), potentially from an observational study. Let $\mathcal{F} = \{f_1, f_2 \cdots f_p\}$ denote the set of all the *characteristics* in $\mathcal{D}$. Consequently, $\mathbf{x}_i = \left[x_i^{(1)}, x_i^{(2)}, \cdots x_i^{(p)}\right] \in [\mathcal{V}_1, \mathcal{V}_2, \cdots \mathcal{V}_p]$ is a vector capturing the values assumed by subject i for each of the characteristics in $\mathcal{F}$. $\mathcal{V}_q$ denotes the set of all possible values that the characteristic f_q can take. In the medical setting, example characteristics include patient's age, BMI, gender, glucose level etc. Let $\mathcal{A}$ denote the set of all possible treatments. $a_i \in \mathcal{A}$ and $y_i \in \mathbb{R}$ represent the *treatment* assigned to subject i and the corresponding *outcome* respectively. We assume that y_i is defined such that higher values indicate better outcomes. For example, the outcome of a patient can be regarded as a wellness improvement score that indicates the effectiveness of the assigned treatment.

It can be much more expensive to determine certain subject characteristics compared to others. For instance, a patient's age can be easily retrieved either from previous records or by asking the patient. On the other hand, determining her glucose level requires more comprehensive testing, and is therefore more expensive in terms of monetary costs, time and effort required both from the patient as well as the clinicians. We assume access to a function $d : \mathcal{F} \rightarrow \mathbb{R}$ which returns the cost of determining any characteristic in $\mathcal{F}$ (assessment cost). The cost associated with a given characteristic $f \in \mathcal{F}$ is assumed to be the same for all the subjects in the population, though the framework can be extended to have patient-specific costs. Analogously, each treatment $a \in \mathcal{A}$ incurs a cost (treatment cost) and we assume access to a function $d' : \mathcal{A} \rightarrow \mathbb{R}$ that returns the cost associated with treatment $a \in \mathcal{A}$.

3.2 Treatment Regimes

A treatment regime is a function that takes as input the characteristics of any given subject $\mathbf{x}$ and maps them to an appropriate treatment $a \in \mathcal{A}$. As discussed, prior studies [29, 23] suggest that decision makers such as doctors and judges who make high stake decisions are more likely to trust, and, therefore employ models that are interpretable. We thus employ *decision lists* to express treatment regimes (see example in Figure 1). A decision list is an ordered list of rules embedded within an if-then-else structure. A treatment regime[1] expressed as a decision list π is a sequence of $L+1$ rules $[r_1, r_2, \cdots, r_{L+1}]$. The last one, r_{L+1}, is a default rule which applies to all those subjects who do not satisfy any of the previous L rules. Each rule r_j (except the default rule) is a tuple of the form (c_j, a_j) where $a_j \in \mathcal{A}$, and c_j represents a *pattern* which is a conjunction of one or more predicates. Each predicate takes the form (f, o, v) where $f \in \mathcal{F}$, $o \in \{=, \neq, \leq, \geq, <, >\}$, and v denotes some value that can be assumed by the characteristic f. Example of such a pattern is "Age $\geq$ 40 $\wedge$ Gender=Female". A subject i is said to satisfy rule j if his/her characteristics $\mathbf{x}_i$ satisfy all the predicates in c_j. Let us formally denote this using an indicator function, *satisfy*$(\mathbf{x}_i, c_j)$ which returns a 1 if $\mathbf{x}_i$ satisfies c_j and 0 otherwise.

The rules in π partition the dataset $\mathcal{D}$ into $L+1$ groups: $\{\mathcal{R}_1, \mathcal{R}_2 \cdots \mathcal{R}_L, \mathcal{R}_{\text{default}}\}$. A group $\mathcal{R}_j$, where $j \in \{1, 2, \cdots L\}$, is comprised of those subjects that satisfy c_j but do not satisfy any of $c_1, c_2, \cdots c_{j-1}$:

$$\mathcal{R}_j = \left\{ \mathbf{x} \in [\mathcal{V}_1 \cdots \mathcal{V}_p] \mid \text{satisfy}(\mathbf{x}, c_j\} \wedge \bigwedge_{t=1}^{j-1} \neg \text{ satisfy}(\mathbf{x}, c_t) \right\}. \tag{1}$$

The treatment assigned to each subject by π is determined by the group that he/she belongs to. For instance, if subject i with characteristics $\mathbf{x}_i$ belongs to group $\mathcal{R}_j$ induced by π i.e., $\mathbf{x}_i \in \mathcal{R}_j$, then subject i will be assigned the corresponding treatment a_j under regime π i.e., $\pi(\mathbf{x}_i) = a_j$.

Similarly, the cost incurred when we assign a treatment to the subject i (*treatment cost*) according to regime π is:

$$\phi(\mathbf{x}_i) = d'(\pi(\mathbf{x}_i)) \tag{2}$$

where the function d', defined in Section 3.1, takes as input a treatment $a \in \mathcal{A}$ and returns its cost.

We can also define the cost incurred in assessing the condition of a subject i (*assessment cost*) as per the regime π. Recall that a subject i belongs to the group $\mathcal{R}_j$ if and only if the subject does not satisfy the conditions $c_1 \cdots c_{j-1}$, but satisfies the condition c_j (Refer Eqn. 1). To reach this conclusion, all the characteristics present in the corresponding antecedents $c_1 \cdots c_j$ must have been measured for subject i and evaluated against the appropriate predicate conditions. This implies that the assessment cost incurred for this subject i is the sum of the costs of all the characteristics that appear in $c_1 \cdots c_j$. If $\mathcal{N}_l$ denotes the set of all the characteristics that appear in $c_1 \cdots c_l$, the assessment cost of the subject i as per the regime π can be written as:

$$\psi(\mathbf{x}_i) = \sum_{l=1}^{L} \left[\mathbb{1}(\mathbf{x}_i \in \mathcal{R}_l) \times \left(\sum_{e \in \mathcal{N}_l} d(e) \right) \right]. \tag{3}$$

where $\mathbb{1}$ denotes an indicator function that returns 1 if the condition within the brackets is true and 0 otherwise.

3.3 Objective Function

We now formulate the objective function for learning a cost-effective treatment regime. We first formalize the no-

[1] We use the terms decision list and treatment regime interchangeably from here on.

tions of expected outcome, assessment, and treatment costs of a treatment regime π with respect to the dataset $\mathcal{D}$.

Expected Outcome Recall that the treatment regime π assigns a subject i with characteristics $\mathbf{x}_i$ to a treatment $\pi(\mathbf{x}_i)$. The quality of regime π is partly determined by the expected outcome when all the subjects in $\mathcal{D}$ are assigned treatments according to regime π. The higher the value of such an expected outcome, the better the quality of π. There is, however, one caveat to computing the value of this expected outcome – we only observe the outcome y_i resulting from assigning $\mathbf{x}_i$ to a_i in the data $\mathcal{D}$, and not any of the counterfactuals. If the regime π assigns a different treatment $a' \neq a_i$ to $\mathbf{x}_i$, we cannot readily determine the corresponding outcome from the data.

The solutions proposed to compute expected outcomes in settings such as ours can be categorized as: adjustment by regression modeling, adjustment by inverse propensity score weighting, and doubly robust estimation. A detailed treatment of each of these approaches is presented by Lunceford et al. [22]. The success of regression-based modeling and inverse weighting depends heavily on the postulated regression model and the postulated propensity score model respectively. In either case, if the postulated models are not identical to the true models, we have biased (inconsistent) estimates of the expected outcome. On the other hand, doubly robust estimation combines the above approaches in such a way that the estimated value of the expected outcome is unbiased as long as one of the postulated models is identical to the true model and there are no unmeasured confounders. The doubly robust estimator for the expected outcome of regime π, denoted by $g_1(\pi)$, can be written as:

$$g_1(\pi) = \frac{1}{N}\sum_{i=1}^{N}\sum_{a\in\mathcal{A}} o(i,a), \text{ where} \tag{4}$$

$$o(i,a) = \left[\frac{\mathbb{1}(a_i = a)}{\hat{\omega}(\mathbf{x}_i, a)}(y_i - \hat{y}(\mathbf{x}_i, a)) + \hat{y}(\mathbf{x}_i, a)\right]\mathbb{1}(\pi(\mathbf{x}_i) = a).$$

$\hat{\omega}(x_i, a)$ denotes the probability that the subject i with characteristics $\mathbf{x}_i$ is assigned to treatment a in the data $\mathcal{D}$. $\hat{\omega}$ represents the propensity score model. In practice, we fit a multinomial logistic regression model on $\mathcal{D}$ to learn this function. Similarly, $\hat{y}(\mathbf{x}_i, a)$ denotes the predicted outcome when a subject characterized by $\mathbf{x}_i$ is assigned to a treatment a. $\hat{y}$ is learned in our experiments by fitting a linear regression model on $\mathcal{D}$ prior to optimizing for the treatment regime. Note that our framework does not impose any constraints on the functional forms of $\hat{y}$ and $\hat{\omega}$ i.e., $\hat{y}$ and $\hat{\omega}$ could be modeled using any suitable technique.

Expected Assessment Cost Recall that there are assessment costs associated with each subject. These costs are governed by the characteristics that will be used in assessing the subject's condition and recommending a treatment. The assessment cost of a subject i treated using the regime π is given in Eqn. 3. The expected assessment cost across the entire population can be computed as:

$$g_2(\pi) = \frac{1}{N}\sum_{i=1}^{N}\psi(\mathbf{x}_i). \tag{5}$$

It is important to ensure that our learning process favors regimes with smaller values of expected assessment cost. Keeping this cost low also ensures that the learned decision list is sparse, which assists with interpretability.

Expected Treatment Cost The treatment cost for a subject i who is assigned treatment using a regime π is given in Eqn. 2. The expected treatment cost across the entire population can be computed as:

$$g_3(\pi) = \frac{1}{N}\sum_{i=1}^{N}\phi(\mathbf{x}_i). \tag{6}$$

The smaller the expected treatment cost of the regime, the more desirable it is in practice. We present the complete objective function below.

Complete Objective We assume access to the following inputs: 1) the observational data $\mathcal{D}$; 2) a set $\mathcal{FP}$ of frequently occurring *patterns* in $\mathcal{D}$. Recall that each pattern corresponds to a conjunction of one or more predicates. An example of such a pattern is "Age $\geq$ 40 $\wedge$ Gender=Female". In practice, such patterns can be obtained by running a frequent pattern mining algorithm such as Apriori [2] on the set $\mathcal{D}$; 3) a set of all possible treatments $\mathcal{A}$.

We define the set of all possible (pattern, treatment) tuples as $\mathcal{L} = \{(c, a)\,|c \in \mathcal{FP}, a \in \mathcal{A}\}$, and $C(\mathcal{L})$ as the set of the permutations of all possible subsets (excluding the null set) of $\mathcal{L}$. An element in $\mathcal{L}$ can be thought of as a rule in a decision list and an element in $C(\mathcal{L})$ can be thought of as a list of rules in a decision list (without the default rule). We then search over all elements in the set $C(\mathcal{L}) \times \mathcal{A}$ to find a regime that maximizes the expected outcome (Eqn. 4) while minimizing the expected assessment (Eqn. 5), and treatment costs (Eqn. 6) all of which are computed over $\mathcal{D}$. Our objective function can be formally written as:

$$\underset{\pi\in C(\mathcal{L})\times\mathcal{A}}{\arg\max}\ \lambda_1 g_1(\pi) - \lambda_2 g_2(\pi) - \lambda_3 g_3(\pi) \tag{7}$$

where g_1, g_2, g_3 are defined in Eqns. 4, 5, 6 respectively, and λ_1 and λ_2 are non-negative weights that scale the relative influence of the terms in the objective.

Theorem 1 *The objective function in Eqn. 7 is NP-hard. (Please see appendix for details.)*

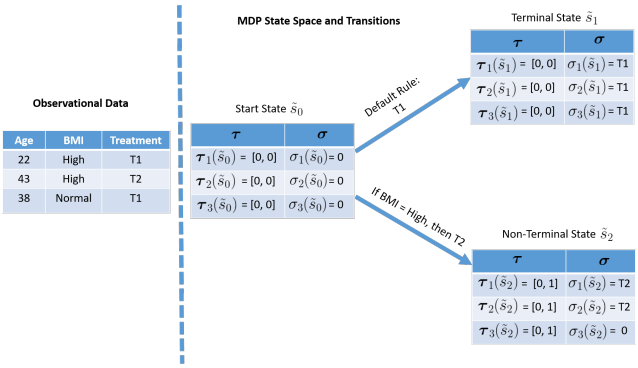

Figure 2: Sample Observational Data and the corresponding Markov Decision Process Representation

Note that NP-hardness is a worst case categorization only; with an efficient search procedure, it is practical to obtain a good approximation on most reasonably-sized datasets.

3.4 Optimizing the Objective

We optimize our objective by modeling it as as a Markov Decision Process (MDP) and then employing Upper Confidence Bound on Trees (UCT) algorithm to find a treatment regime which maximizes Eqn. 7. We also propose and leverage customized checks for guiding the exploration of the UCT algorithm and pruning the search space effectively.

Markov Decision Process Formulation. Our goal is to find a sequence of rules that maximize the objective function in Eqn. 7. To this end, we formulate a fully observable MDP such that the optimal policy of the posited formulation provides a solution to our optimization problem.

A fully observable MDP is characterized by a tuple $(\mathbf{S}, \mathbf{A}, \mathbf{T}, \mathbf{R})$ where $\mathbf{S}$ denotes the set of all possible states, $\mathbf{A}$ denotes the set of all possible actions, $\mathbf{T}$ and $\mathbf{R}$ represent the transition and reward functions respectively. Below we define each of these in the context of our problem. Figure 2 shows a snapshot of the state space and transitions for a small dataset.

State Space. Conceptually, each state in our state space captures the effect of some partial or fully constructed decision list. To illustrate, let us consider a partial decision list with just one rule "if Age $\geq$ 40 $\wedge$ Gender = Female, then T1". This partial list enforces that: (i) all those subjects that satisfy the condition of the rule are assigned treatment T1, and (ii) age and gender characteristics will be required in determining treatments for all the subjects in the population.

To capture such information, we represent a state $\tilde{s} \in \boldsymbol{S}$ by a list of tuples $[(\boldsymbol{\tau}_1(\tilde{s}), \sigma_1(\tilde{s})), \cdots (\boldsymbol{\tau}_N(\tilde{s}), \sigma_N(\tilde{s}))]$ where each tuple corresponds to a subject in $\mathcal{D}$. $\boldsymbol{\tau}_i(\tilde{s})$ is a binary vector of length p defined such that $\tau_i^{(j)}(\tilde{s}) = 1$ if the characteristic j will be required for determining subject i's treatment in state $\tilde{s}$, and 0 otherwise. Further, $\sigma_i(\tilde{s})$ captures the treatment assigned to subject i in state $\tilde{s}$. If no treatment has been assigned to i, then $\sigma_i(\tilde{s}) = 0$.

Note that we have a single start state $\tilde{s}_0$ which corresponds to an empty decision list. $\boldsymbol{\tau}_i(\tilde{s}_0)$ is a vector of 0s, and $\sigma_i(\tilde{s}_0) = 0$ for all subjects i in $\mathcal{D}$ indicating that no treatments have been assigned to any subject, and no characteristics were deemed as requirements for assigning treatments. Furthermore, a state $\tilde{s}$ is regarded as a terminal state if for all i, $\sigma_i(\tilde{s})$ is non-zero indicating that treatments have been assigned to all the subjects.

Actions. Each action can take one of the following forms: 1) a rule $r \in \mathcal{L}$, which is a tuple of the form (pattern, treatment). Eg., (Age$\geq$40 $\wedge$ Gender=Female, T1). This specifies that subjects who obey conditions in the pattern are prescribed the corresponding treatment. Such action leads to a non-terminal state. 2) a treatment $a \in \mathcal{A}$, which corresponds to the default rule leading to a terminal state.

Transition and Reward Functions. We have a deterministic transition function which ensures that taking an action $\tilde{a} = (\tilde{c}, \tilde{t})$ from state $\tilde{s}$ will always lead deterministically to a state $\tilde{s}'$. Let U denote the set of all those subjects i for which treatments have already been assigned in state $\tilde{s}$ i.e., $\sigma_i(\tilde{s}) \neq 0$ and let U^c denote the set of all those subjects who have not been assigned treatment in the state $\tilde{s}$. Let U' denote the set of all those subjects i that do not belong to the set U and that satisfy the condition $\tilde{c}$ of action $\tilde{a}$. Let Q denote the set of all those characteristics in $\mathcal{F}$ that are present in the condition $\tilde{c}$ of action $\tilde{a}$. If action $\tilde{a}$ corresponds to a default rule, then $Q = \emptyset$ and $U' = U^c$. With this notation in place, the new state $\tilde{s}'$ can be characterized as follows: 1) $\tau_i^{(j)}(\tilde{s}') = \tau_i^{(j)}(\tilde{s})$ and $\sigma_i(\tilde{s}') = \sigma_i(\tilde{s})$ for all $i \in U$, $j \in \mathcal{F}$; 2) $\tau_i^{(j)}(\tilde{s}') = 1$ for all $i \in U^c$, $j \in Q$; 3) $\sigma_i(\tilde{s}') = \tilde{t}$ for all $i \in U'$.

The immediate reward obtained when we reach $\tilde{s}'$ by taking action $\tilde{a} = (\tilde{c}, \tilde{t})$ from the state $\tilde{s}$ can be written as:

$$\frac{\lambda_1}{N} \sum_{i \in U'} o(i, \tilde{t}) - \frac{\lambda_2}{N} \sum_{i \in U^c, j \in Q} d(j) - \frac{\lambda_3}{N} \sum_{i \in U'} d'(\tilde{t})$$

where o is defined in Eqn. 4, d and d' are cost functions for characteristics and treatments respectively (see Section 3.1).

UCT with Customized Pruning. The basic idea behind the Upper Confidence Bound on Trees (UCT) [12] algorithm is to iteratively construct a search tree for some predetermined number of iterations. At the end of this procedure, the best performing policy or sequence of actions is returned as the output. Each node in the search tree corresponds to a state in the MDP state space and the links in the tree correspond to the actions. UCT employs the UCB-1 metric [6] for navigating through the search space.

We employ a UCT-based algorithm for finding the optimal

	Bail Dataset	Asthma Dataset
# of Data Points	86152	60048
Characteristics & Costs	age, gender, previous offenses, prior arrests, current charge, SSN (cost = 1) marital status, kids, owns house, pays rent addresses in past years (cost = 2) mental illness, drug tests (cost = 6)	age, gender, BMI, BP, short breath, temperature, cough, chest pain, wheezing, past allergies, asthma history, family history, has insurance (cost 1) peak flow test (cost = 2) spirometry test (cost = 4) methacholine test (cost = 6)
Treatments & Costs	release on personal recognizance (cost = 20) release on conditions/bond (cost = 40)	quick relief (cost = 10) controller drugs (cost = 15)
Outcomes & Scores	no risk (score = 100), failure to appear (score = 66) non-violent crime (score = 33) violent crime (score = 0)	no asthma attack for $\geq$ 4 months (score = 100) no asthma attack for 2 months (score = 66) no asthma attack for 1 month (score = 33) asthma attack in less than 2 weeks (score = 0)

Table 1: Summary of datasets.

policy of our MDP formulation, though we leverage customized checks to further guide the exploration process and prune the search space. Recall that each non-terminal state in our state space corresponds to a partial decision list. We exploit the fact that we can upper bound the value of the objective for any given partial decision list. The upper bound on the objective for any given non-terminal state $\tilde{s}$ can be computed by approximating the reward as follows: 1) all the subjects who have not been assigned treatments will get the best possible treatments without incurring any treatment cost 2) no additional assessments are required by any subject (and hence no additional assessment costs levied) in the population. The upper bound on the incremental reward is thus:

$$\text{upper bound}(U^c) = \frac{\lambda_1}{N} \sum_{i \in U^c} \max_t o(i,t).$$

During the execution of UCT procedure, whenever there is a choice to be made about which action needs to be taken, we employ checks based on the upper bound of the objective value of the resulting state. Consider a scenario in which the UCT procedure is currently in state $\tilde{s}$ and needs to choose an action. For each possible action $\tilde{a}$ (that does not correspond to a default rule) from state $\tilde{s}$, we determine the upper bound on the objective value of the resulting state $\tilde{s}'$. If this value is less than either the highest value encountered previously for a complete rule list, or the objective value corresponding to the best default action from the state $\tilde{s}$, then we *block* the action $\tilde{a}$ from the state $\tilde{s}$. This state is provably suboptimal. Note that we can compute exact values of the objective function if the action is a default rule because the corresponding decision list is fully constructed.

4 Experimental Evaluation

Here, we discuss the detailed experimental evaluation of our framework. First we analyze the outcomes obtained and costs incurred when recommending treatments using our approach. Then, we present an ablation study which explores the contributions of each of the terms in our objective, followed by an analysis on real world data.

Dataset Description. Our first dataset consists of information pertaining to the **bail** decisions [37, 15] of about 86K defendants collected from various state courts in the U.S. between 1990-2009 (Table 1). It captures information about various characteristics for each of the 86K defendants. The decisions made by judges in each of these cases (release without/with conditions) and the corresponding outcomes (e.g., if a defendant committed another crime when out on bail) are also available. We assigned costs to characteristics, and treatments based on discussions with subject matter experts. The characteristics that were harder to obtain were assigned higher costs compared to the ones that were readily available. Similarly, the treatment that placed a higher burden on the defendant (release on condition) was assigned a higher cost. We assigned each outcome a numerical score and higher score indicates a better outcome. Thus, undesirable scenarios (e.g., violent crime when released on bail) received lower scores.

Our second dataset (Table 1) captures details of about 60K **asthma** patients collected by a web-based electronic health record company [15]. For each of these 60K patients, various attributes such as demographics, symptoms, past health history, test results have been recorded. Each patient was prescribed either quick relief medications or long term controller drugs. Further, the outcomes in the form of time to the next asthma attack (after the treatment began) were recorded. The longer this interval, the better the outcome, and the higher the outcome score. We assigned costs to characteristics, and treatments based on the inconvenience (physical/emotional/monetary) they caused patients.

Baselines. We compared our framework to the following state-of-the-art treatment recommendation approaches: 1) Outcome Weighted Learning (OWL) [42] 2) Modified Covariate Approach (MCA) [31] 3) Interpretable and Parsimonious Treatment Regime Learning (IPTL) [39]. OWL addresses the problem of treatment recommendation by formulating it as a weighted classification problem where each subject is weighted proportional to his/her outcome value. MCA generates modified covariates which capture the interactions between characteristics of subjects and

	Bail Dataset					Asthma Dataset				
	Avg. Outcome	Avg. Assess Cost	Avg. Treat Cost	Avg. # of Characs.	List Len	Avg. Outcome	Avg. Assess Cost	Avg. Treat Cost	Avg. # of Characs.	List Len
CITR	79.2	8.88	31.09	6.38	7	74.38	13.87	11.81	7.23	6
IPTL	77.6	14.53	35.23	8.57	9	71.88	18.58	11.83	7.87	8
MCA	73.4	19.03	35.48	12.03	-	70.32	19.53	12.01	10.23	-
OWL (Gaussian)	72.9	28	35.18	13	-	71.02	25	12.38	16	-
OWL (Linear)	71.3	28	34.23	13	-	71.02	25	12.38	16	-
CITR - No Treat	80.5	8.93	34.48	7.57	7	77.39	14.02	12.87	7.38	7
CITR - No Assess	81.3	13.83	32.02	9.86	10	78.32	18.28	12.02	8.97	9
CITR - Outcome	81.7	13.98	34.49	10.38	10	79.37	18.28	12.88	9.21	9
Human	69.37	-	33.39	-	-	68.32	-	12.28	-	-

Table 2: Results for Treatment Regimes. Our approach: CITR; Baselines: IPTL, MCA, OWL; Ablations of our approach: CITR - No Treat, CITR - No Assess, CITR - Outcome; Human refers to the setting where judges and doctors assigned treatments.

treatments and then uses these modified covariates to fit a model for predicting the outcomes. The treatments of subjects are then determined based on the values of the predicted outcomes. The IPTL framework produces interpretable decision lists which map subject characteristics to treatments such that the resulting outcomes are maximized. IPTL, however, does not account for assessment or treatment costs. While MCA and IPTL minimize the number of characteristics/covariates required for deciding the treatment of any given subject, OWL utilizes all the characteristics available in the data when assigning treatments.

Experimental Setting. The objective function that we proposed in Eqn. 7 has three parameters $\lambda_1, \lambda_2,$ and λ_3. These parameters could either be specified by an end-user or learned using a validation set. We set aside 5% of each of our datasets as a validation set to estimate these parameters. We automatically searched the parameter space to find a set of parameters that produced a decision list with the maximum average outcome on the validation set (discussed in detail later) and satisfied some simple constraints such as: 1) average assessment cost ≤ 4 on both the datasets 2) average treatment cost ≤ 30 for the bail data; average treatment cost ≤ 12 for the asthma data. We then used a coordinate ascent strategy to search the parameter space and update each parameter λ_j while holding the other two parameters constant. The values of each of these parameters were chosen via a binary search on the interval $(0, 1000)$. We ran the UCT procedure for 50K iterations to generate our decision list. We used both Gaussian and linear kernels for OWL and employed the tuning strategy discussed in Zhao et. al. [42]. In case of IPTL, we set the parameter that limits the number of the rules in the treatment regime to 20. We evaluated the performance of our model and other baselines using 10 fold cross validation.

4.1 Quantitative Evaluation

We analyzed the performance of our approach CITR (Cost-effective, Interpretable Treatment Regimes) on various aspects such as outcomes obtained, costs incurred, and intelligibility. We computed the following metrics:

Avg. Outcome Recall that a treatment regime assigns a treatment to every subject in the population. We used the prediction model $\hat{y}$ (defined in Section 3.3) to obtain an outcome score given the characteristics of the subject and the treatment assigned (we used ground truth outcome scores whenever they were available in the data). We computed the average outcome score of all the subjects in the population.

Avg. Assess Cost We determined assessment costs incurred by each subject based on what characteristics were used to determine their treatment. We then averaged all such per-subject assessment costs to obtain the average assessment cost.

Avg. # of Characs We determined the number of characteristics that are used when assigning a treatment to each subject in the population and then computed the average of these numbers.

Avg. Treat Cost We computed the average of the treatment costs incurred by all the subjects in the population.

List Len Our approach CITR and the baseline IPTL express treatment regimes as decision lists. In order to compare the complexity of the resulting decision lists, we computed the number of rules in each of these lists.

While higher values of average outcome are preferred, lower values on all of the other metrics are desirable.

Results. Table 2 (top panel) presents the values of the metrics computed for our approach as well as the baselines. It can be seen that the treatment regimes produced by our approach result in better average outcomes with lower costs across both the datasets. While IPTL and MCA do not explicitly reduce costs, they do minimize the number of characteristics required for determining treatment of any given subject. Yet, our approach produces regimes with the lowest values of the average number of characteristics (Avg. # of Characs). Our approach also produces more concise lists with fewer rules compared to the baselines. While the treatment costs of all the baselines are similar, there is some variation in the average assessment costs and the outcomes. IPTL turns out to be the best performing baseline in terms of the average outcome, average assess-

ment costs, and average number of characteristics. The last line of Table 2 shows the average outcomes and the average treatment costs computed empirically on the observational data. Both of our datasets are comprised of decisions made by human experts. It is interesting to note that the regimes learned by algorithmic approaches possibly result in better outcomes compared to the decisions made by human experts on both the datasets.

4.1.1 Ablation Study

We also analyzed the effect of various terms of our objective function on the outcomes, and the costs incurred. To this end, we experimented with three different ablations of our approach: 1) *CITR - No Treat*, obtained by excluding the term corresponding to the expected treatment cost in our objective ($g_3(\pi)$ in Eqn. 7). 2) *CITR - No Assess*, obtained by excluding the expected assessment cost term in our objective ($g_2(\pi)$) in Eqn. 7) 3) *CITR - Outcome*, obtained by excluding both assessment and treatment cost terms from our objective.

Table 2 (second panel) shows the values of the metrics discussed earlier in this section for various ablations of our model. Naturally, removing the treatment cost term increases the average treatment cost on both datasets. Furthermore, removing the assessment cost term of the objective results in regimes with much higher assessment costs (8.88 vs. 13.83 on bail data; 13.87 vs. 18.28 on asthma data). The length of the list also increases for both datasets when we exclude the assessment cost term. These results demonstrate that each term in our objective function is crucial to producing cost-effective interpretable regimes.

4.2 Qualitative Analysis

The treatment regimes produced by our approach on asthma and bail datasets are shown in Figures 1 and 3 respectively.

It can be seen in Figure 3 that the methacholine test, which is more expensive, appears at the end of the regime. This ensures that only a small fraction of the population (8.23%) is burdened by its cost. Furthermore, it turns out that though the spirometry test is slightly expensive compared to patient demographics and symptoms, it would be harder to determine the treatment for a patient without this test. This aligns with research on asthma treatment recommendations [26, 5]. It is interesting to note that the regime not only accounts for test results on spirometry and peak flow but also evaluates whether the patient has a previous history of asthma or respiratory issues. If the test results are positive and the patient has no previous history of asthma or respiratory disorders, then the patient is recommended quick relief drugs. On the other hand, if the test results are positive and the patient suffered previous asthma or respiratory issues, then controller drugs are recommended.

In case of the bail dataset, the constructed regime is able to achieve good outcomes without even using the most expensive characteristics such as mental illness tests and drug tests. Personal information characteristics, which are slightly more expensive than defendant demographics and prior criminal history, appear only towards the end of the list and these checks apply only to 21.23% of the population. Note that the regime uses defendants' criminal history as well as personal and demographic information to make treatment recommendations. For instance, females with minor current charges (such as driving offenses) and no prior criminal record are typically released on bail without conditions such as bonds or checking in with the police. On the other hand, defendants who have committed crimes earlier are only granted conditional bail.

If Gender=F **and** Current-Charge =Minor **and** Prev-Offense=None **then** RP

Else if Prev-Offense=Yes **and** Prior-Arrest =Yes **then** RC

Else if Current-Charge =Misdemeanor **and** Age$\leq$ 30 **then** RC

Else if Age$\geq$ 50 **and** Prior-Arrest=No, **then** RP

Else if Marital-Status=Single **and** Pays-Rent =No **and** Current-Charge =Misd. **then** RC

Else if Addresses-Past-Yr$\geq$ 5 **then** RC

Else RP

Figure 3: Treatment regime for bail data; RP refers to milder form of treatment: release on personal recognizance, and RC is release on condition which is comparatively harsher.

5 Conclusions

In this work, we proposed a framework for learning cost-effective, interpretable treatment regimes from observational data. To the best of our knowledge, this is the first solution to the problem at hand that addresses all of the following aspects: 1) maximizing the outcomes 2) minimizing the treatment costs, and costs associated with gathering information required to determine the treatment 3) expressing regimes using an interpretable model. We modeled the problem of learning a treatment regime as a MDP and employed a variant of UCT which prunes the search space using customized checks. We demonstrated the effectiveness of our framework on real world data from criminal justice and health care domains.

6 Acknowledgments

H. Lakkaraju is funded by a Robert Bosch Stanford Graduate Fellowship. C. Rudin was partially supported by Adobe. The authors would like to thank Aditya Grover, and the anonymous reviewers for their insightful comments and feedback.

References

[1] Eva-Maria Abulesz and Gerasimos Lyberatos. Novel approach for determining optimal treatment regimen for cancer cemotherapy. *International journal of systems science*, 19(8):1483–1497, 1988.

[2] Rakesh Agrawal, Ramakrishnan Srikant, et al. Fast algorithms for mining association rules.

[3] James O Berger, Xiaojing Wang, and Lei Shen. A bayesian approach to subgroup identification. *Journal of biopharmaceutical statistics*, 24(1):110–129, 2014.

[4] Jacob Bien and Robert Tibshirani. Classification by set cover: The prototype vector machine. *arXiv preprint arXiv:0908.2284*, 2009.

[5] Louis-Philippe Boulet, Marie-Ève Boulay, Guylaine Gauthier, Livia Battisti, Valérie Chabot, Marie-France Beauchesne, Denis Villeneuve, and Patricia Côté. Benefits of an asthma education program provided at primary care sites on asthma outcomes. *Respiratory medicine*, 109(8):991–1000, 2015.

[6] Cameron B Browne, Edward Powley, Daniel Whitehouse, Simon M Lucas, Peter I Cowling, Philipp Rohlfshagen, Stephen Tavener, Diego Perez, Spyridon Samothrakis, and Simon Colton. A survey of monte carlo tree search methods. *IEEE Transactions on Computational Intelligence and AI in Games*, 4(1):1–43, 2012.

[7] Ailin Fan, Wenbin Lu, Rui Song, et al. Sequential advantage selection for optimal treatment regime. *The Annals of Applied Statistics*, 10(1):32–53, 2016.

[8] Jared C Foster, Jeremy MG Taylor, and Stephen J Ruberg. Subgroup identification from randomized clinical trial data. *Statistics in medicine*, 30(24):2867–2880, 2011.

[9] Kosuke Imai, Marc Ratkovic, et al. Estimating treatment effect heterogeneity in randomized program evaluation. *The Annals of Applied Statistics*, 7(1):443–470, 2013.

[10] Shihao Ji and Lawrence Carin. Cost-sensitive feature acquisition and classification. *Pattern Recognition*, 40(5):1474–1485, 2007.

[11] Been Kim, Cynthia Rudin, and Julie A Shah. The bayesian case model: A generative approach for case-based reasoning and prototype classification. In *Advances in Neural Information Processing Systems*, pages 1952–1960, 2014.

[12] Levente Kocsis and Csaba Szepesvári. Bandit based monte-carlo planning. In *European conference on machine learning*, pages 282–293. Springer, 2006.

[13] EB Laber and YQ Zhao. Tree-based methods for individualized treatment regimes. *Biometrika*, 102(3):501–514, 2015.

[14] Eric B Laber, Daniel J Lizotte, Min Qian, William E Pelham, and Susan A Murphy. Dynamic treatment regimes: Technical challenges and applications. *Electronic journal of statistics*, 8(1):1225, 2014.

[15] Himabindu Lakkaraju, Stephen H Bach, and Jure Leskovec. Interpretable decision sets: A joint framework for description and prediction. In *KDD*, 2016.

[16] Himabindu Lakkaraju and Jure Leskovec. Confusions over time: An interpretable bayesian model to characterize trends in decision making. In *Advances in Neural Information Processing Systems (NIPS)*, 2016.

[17] Himabindu Lakkaraju, Jure Leskovec, Jon Kleinberg, and Sendhil Mullainathan. A bayesian framework for modeling human evaluations. In *SIAM SDM*, 2015.

[18] Benjamin Letham, Cynthia Rudin, Tyler H McCormick, David Madigan, et al. Interpretable classifiers using rules and bayesian analysis: Building a better stroke prediction model. *The Annals of Applied Statistics*, 9(3):1350–1371, 2015.

[19] Wei-Yin Loh, Xu He, and Michael Man. A regression tree approach to identifying subgroups with differential treatment effects. *Statistics in medicine*, 34(11):1818–1833, 2015.

[20] Susan Lomax and Sunil Vadera. A survey of cost-sensitive decision tree induction algorithms. *ACM Computing Surveys (CSUR)*, 45(2):16, 2013.

[21] Yin Lou, Rich Caruana, and Johannes Gehrke. Intelligible models for classification and regression. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 150–158. ACM, 2012.

[22] Jared K Lunceford and Marie Davidian. Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. *Statistics in medicine*, 23(19):2937–2960, 2004.

[23] Douglas B Marlowe, David S Festinger, Karen L Dugosh, Kathleen M Benasutti, Gloria Fox, and Jason R Croft. Adaptive programming improves outcomes in drug court an experimental trial. *Criminal justice and behavior*, 39(4):514–532, 2012.

[24] Erica EM Moodie, Bibhas Chakraborty, and Michael S Kramer. Q-learning for estimating optimal dynamic treatment rules from observational data. *Canadian Journal of Statistics*, 40(4):629–645, 2012.

[25] Erica EM Moodie, Nema Dean, and Yue Ru Sun. Q-learning: Flexible learning about useful utilities. *Statistics in Biosciences*, 6(2):223–243, 2014.

[26] Jorge Pereira, Priscilla Porto-Figueira, Carina Cavaco, Khushman Taunk, Srikanth Rapole, Rahul Dhakne, Hampapathalu Nagarajaram, and José S Câmara. Breath analysis as a potential and non-invasive frontier in disease diagnosis: an overview. *Metabolites*, 5(1):3–55, 2015.

[27] Min Qian and Susan A Murphy. Performance guarantees for individualized treatment rules. *Annals of statistics*, 39(2):1180, 2011.

[28] James M Robins. Correcting for non-compliance in randomized trials using structural nested mean models. *Communications in Statistics-Theory and methods*, 23(8):2379–2412, 1994.

[29] Richard N Shiffman. Representation of clinical practice guidelines in conventional and augmented decision tables. *Journal of the American Medical Informatics Association*, 4(5):382–393, 1997.

[30] Xiaogang Su, Chih-Ling Tsai, Hansheng Wang, David M Nickerson, and Bogong Li. Subgroup analysis via recursive partitioning. *Journal of Machine Learning Research*, 10(Feb):141–158, 2009.

[31] Lu Tian, Ash A Alizadeh, Andrew J Gentles, and Robert Tibshirani. A simple method for estimating interactions between a treatment and a large number of covariates. *Journal of the American Statistical Association*, 109(508):1517–1532, 2014.

[32] Stijn Vansteelandt, Marshall Joffe, et al. Structural nested models and g-estimation: The partially realized promise. *Statistical Science*, 29(4):707–731, 2014.

[33] Michael P Wallace and Erica EM Moodie. Personalizing medicine: a review of adaptive treatment strategies. *Pharmacoepidemiology and drug safety*, 23(6):580–585, 2014.

[34] Fulton Wang and Cynthia Rudin. Causal falling rule lists. *CoRR*, abs/1510.05189, 2015.

[35] Fulton Wang and Cynthia Rudin. Falling rule lists. In *Proceedings of Artificial Intelligence and Statistics (AISTATS)*, 2015.

[36] Tong Wang, Cynthia Rudin, Finale Doshi, Yimin Liu, Erica Klampfl, and Perry MacNeille. Bayesian or's of and's for interpretable classification with application to context aware recommender systems. In *ICDM*, 2016.

[37] Marian R Williams. From bail to jail: The effect of jail capacity on bail decisions. *American Journal of Criminal Justice*, 41(3):484–497, 2016.

[38] Baqun Zhang, Anastasios A Tsiatis, Eric B Laber, and Marie Davidian. A robust method for estimating optimal treatment regimes. *Biometrics*, 68(4):1010–1018, 2012.

[39] Yichi Zhang, Eric B. Laber, Anastasios Tsiatis, and Marie Davidian. Using decision lists to construct interpretable and parsimonious treatment regimes. *Biometrics*, 71(4):895–904, 2015.

[40] Yichi Zhang, Eric B Laber, Anastasios Tsiatis, and Marie Davidian. Interpretable dynamic treatment regimes. *arXiv preprint arXiv:1606.01472*, 2016.

[41] Ying-Qi Zhao, Donglin Zeng, Eric B Laber, Rui Song, Ming Yuan, and Michael Rene Kosorok. Doubly robust learning for estimating individualized treatment with censored data. *Biometrika*, 102(1):151–168, 2015.

[42] Yingqi Zhao, Donglin Zeng, A John Rush, and Michael R Kosorok. Estimating individualized treatment rules using outcome weighted learning. *Journal of the American Statistical Association*, 107(499):1106–1118, 2012.

[43] Yufan Zhao, Michael R Kosorok, and Donglin Zeng. Reinforcement learning design for cancer clinical trials. *Statistics in medicine*, 28(26):3294–3315, 2009.