Methods for Interpretable Treatment Regimes

John Sperger

University of North Carolina at Chapel Hill

May 4th 2018
Why Interpretable Methods?

- Doctors and other decision makers are more likely to trust results they can understand. Patients may be more willing to comply with treatment if they understand why it was given to them.
- Hypothesis Generation and Exploratory Data Analysis
- Auditability / reporting requirements

Source: https://xkcd.com/1838/
A Decision List is a sequence of if-then-else rules

In mathematical logic, an atomic formula is a boolean formula with only a single variable and no logical connectives. Every atomic formula $x$ has two literals associated with it: itself, $x$, and its negation $\neg x$. A term is a conjunction of literals, for example $(x_1 \land x_2)$ or $(x_1 \land \neg x_2)$. [3]

A group $R_j$ is made of those subjects that satisfy the term $t_j$ but not terms $t_1, \ldots, t_{j-1}$, except for $R_{L+1}$ which is comprised of all subjects who don’t satisfy any of the first $L$ terms

$$R_j = \left\{ \mathbf{x} : \text{satisfy}(\mathbf{x}, t_j) \land \neg \text{satisfy}(\mathbf{x}, t_k) \right\}_{k=1}^{j-1}$$
The observed data are $n$ i.i.d. trajectories $(x, a, y)$ and we’re interested in a treatment rule $\pi : X \to A$

- $x \in X \subseteq \mathbb{R}^p$ a vector of $p$ covariates for subject $i$ measured at the beginning of the stage
- $a \in A$ action. For a binary treatment $A = \{-1, 1\}$
- $y \in \mathbb{R}$ is the outcome for patient defined so that higher values correspond to better outcomes
Define the state-action value function $Q(x, a) = \mathbb{E}[Y|X = x, A = a]$.

$Q$-learning with policy search involves constructing an estimator $\hat{Q}$ for the $Q$-function and then search for a $\hat{\pi}$ which maximizes the value of the estimated $Q$-function.

Zhang et al.: Model $Q$ with kernel ridge regression. Use a greedy optimization procedure which turns out to be consistent for the global maximizer. [6]

Lakkaraju & Rudin: Doubly robust estimation for $Q$. Represent fitting a list as an MDP where states are (partial) decision lists. Then use Monte Carlo Tree Search methods to find the best list-based treatment regime. [2]
Expected Outcome:

\[ g_1(\pi) = \frac{1}{n} \sum_{i=1}^{n} \sum_{a \in A} \left[ \frac{I(a_i = a)}{\hat{\omega}(x_i, a)} (y_i - \hat{Q}(x_i, a)) + \hat{Q}(x_i, a) \right] I(\pi(x_i) = a) \]

Expected Assessment Cost: \( g_2(\pi) = \frac{1}{n} \sum_{i=1}^{n} \psi(x_i, \pi) \)

Ordered so that if \( x_i \) belongs to group \( R_i \) and \( x_j \) belongs to group \( R_j \) where \( i < j \) then \( \psi(x_i) \leq \psi(x_j) \)

Expected Treatment Cost: \( g_3(\pi) = \frac{1}{n} \sum_{i=1}^{n} \phi(x_i, \pi) \)
Finally, L&R define the set $FP$ of frequent patterns in the data by running a frequent pattern mining algorithm on the data. The resulting space of policies $FP \times A$ is a subset of the overall space of list-based policies $\Pi$.

Letting $\lambda_1, \lambda_2, \lambda_3$ be non-negative weights the objective function is

$$\pi^{opt} = \arg\max_{\pi \in FP \times A} \lambda_1 g_1(\pi) - \lambda_2 g_2(\pi) - \lambda_3 g_3(\pi)$$
States: partial decision lists (non-terminal states) or fully fitted decision lists (terminal states)

Actions: adding a rule to a decision list

Transitions: The environment’s transition model is deterministic

Reward:
\[
\frac{\lambda_1}{n} \sum_{i \in D'} o(x_i, \hat{a}) - \frac{\lambda_2}{n} \sum_{i \in D^c, j \in T} \psi(x_i, \hat{r}_j) - \frac{\lambda_3}{n} \sum_{i \in D'} \phi(x_i, \hat{a})
\]

$T$: new covariates in $\hat{r}$ which haven’t been previously assessed

$D$: set of subjects whose treatments have already been determined in state $s$

$D^c$: all subjects whose treatment assignments haven’t been determined yet

$D' \subseteq D^c$: are the subjects assigned a treatment in the new rule
Monte-Carlo Tree Search Overview [4]

Repeat while time remains

Selection → Expansion → Simulation → Backup

Tree Policy

Rollout Policy
To choose a node in the selection step, the algorithm picks the node $s$ based on the average empirical value of being in that node and a function of many times that node has been selected $N_s$ out of the total number of iterations so far $N$.

$$\text{UCB1 : } \bar{V}_s + \sqrt{\frac{2 \log N}{N_s}}$$

Over time the probability of choosing the wrong action at the initial node goes to zero [1]

L&R prune the search space by calculating an upper bound. This upper bound is calculated by assuming all the unassigned subjects get the best possible treatment without incurring any assessment or treatment costs.

$$\text{Upper Bound}(D^C) = \frac{\lambda_1}{n} \sum_{i \in D^C} \max_a o(i, a)$$
$Q$-learning Lists

- Start with favorite approach to estimate the $Q$-function. The authors chose kernel ridge regression.
- Determine the estimated optimal decision rule from your estimated $Q$-function $\hat{\pi}^Q = \arg\max_{a \in A} \hat{Q}(x, a)$
- Fit a decision list in a greedy way one clause at a time by finding $R$ and $a$

$$\text{If } x \in R \text{ then } a$$
$$\text{else if } x \in \mathcal{X} \text{ then } \hat{\pi}^Q(x)$$
Fitting the first rule

Let $V(R) \in \{0, 1, 2\}$ be the number of variables used to define $R$, and tuning parameters $\eta, \zeta > 0$. Then the rule and action $(\hat{R}_1, \hat{a}_1)$ that define the first rule in the decision list.

$$\hat{R}_1, \hat{a}_1 = \arg\max_{R,a} \frac{1}{n} \sum_{i=1}^{n} \left[ I(x_i \in R)\hat{Q}(x_i, a) + I(x_i \notin R)\hat{Q}(x, \hat{\pi}_Q(x_i)) \right]$$

$$+ \zeta \left( \frac{1}{n} \sum_{i=1}^{n} I(x_i \in R) \right) + \eta (2 - V(R))$$

(2)
Additional rules

If $x \in \hat{R}_1$ then $\hat{a}_1$

else if $x \in R$ then $a$

else if $x \in D^c$ then $\hat{\pi}^Q(x)$

The first term is independent of $R$ and $a$ and so can be dropped during the optimization procedure. This leads to the general form of the when estimating the $t$-th rule of the decision list

$$
\hat{R}_t, \hat{a}_t = \arg\max_{R,a} \frac{1}{n} \sum_{i=1}^{n} I(x_i \in R, x_i \in D^c) \hat{Q}(x_i, a) + \frac{1}{n} \sum_{i=1}^{n} I(x_i \notin R, x_i \in D^c) \hat{Q}(x, \hat{\pi}^Q(x_i)) + \zeta \left( \frac{1}{n} \sum_{i=1}^{n} I(x_i \in R, x_i \in D^c) \right) + \eta (2 - V(R))
$$
Method Comparison

- Both methods converge to the optimal list-based representation as $n \to \infty$ [1]
- Zhang et al.’s is more computationally efficient. In their paper they extend their approach to the setting where there are $T < \infty$ decision stages.
- Lakkaraju & Rudin’s approach could be extended to a multi-stage setting with $Q$-learning in theory. Computational complexity may be a problem as $T$ increases, but many precision medicine/clinical trial applications have few enough stages that it’s probably feasible.
- Lakkaraju & Rudin’s approach has multiple pre-processing steps which are all sensible but the statistical properties of the pipeline aren’t clear
$R \sim N(Q_0, 1)$ where $Q_0 = 1 + 2X_1 + X_2 + .5X_3 + .442(1 - X_1 - X_2)A$ [7]
Take-away Message

- There’s interesting work going on with interpretable methods.

- You should try using them even if the model you use for treatment assignment is black-box (and it works for classification too!)

- Interpretable methods are only as good as the model they are based on. Model selection and validation are still necessary for reinforcement learning [5]
A survey of monte carlo tree search methods.  

Himabindu Lakkaraju and Cynthia Rudin.  
Learning cost-effective and interpretable treatment regimes.  

Ronald Rivest.  
Learning Decision Lists.  
Richard S Sutton and Andrew G Barto. 

Jeremy MG Taylor, Wenting Cheng, and Jared C Foster. 
Reader reaction to “a robust method for estimating optimal treatment regimes” by zhang et al.(2012). 

Yichi Zhang, Eric B. Laber, Marie Davidian, and Anastasios A. Tsiatis. 
Estimation of optimal treatment regimes using lists. 
### L & R Sim Results

#### Bail Dataset

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>CITR</td>
<td>79.2</td>
<td>8.88</td>
<td>31.09</td>
<td>6.38</td>
<td>7</td>
</tr>
<tr>
<td>IPTL</td>
<td>77.6</td>
<td>14.53</td>
<td>35.23</td>
<td>8.57</td>
<td>9</td>
</tr>
<tr>
<td>MCA</td>
<td>73.4</td>
<td>19.03</td>
<td>35.48</td>
<td>12.03</td>
<td>-</td>
</tr>
<tr>
<td>OWL (Gaussian)</td>
<td>72.9</td>
<td>28</td>
<td>35.18</td>
<td>13</td>
<td>-</td>
</tr>
<tr>
<td>OWL (Linear)</td>
<td>71.3</td>
<td>28</td>
<td>34.23</td>
<td>13</td>
<td>-</td>
</tr>
<tr>
<td>CITR - No Treat</td>
<td>80.5</td>
<td>8.93</td>
<td>34.48</td>
<td>7.57</td>
<td>7</td>
</tr>
<tr>
<td>CITR - No Assess</td>
<td>81.3</td>
<td>13.83</td>
<td>32.02</td>
<td>9.86</td>
<td>10</td>
</tr>
<tr>
<td>CITR - Outcome</td>
<td>81.7</td>
<td>13.98</td>
<td>34.49</td>
<td>10.38</td>
<td>10</td>
</tr>
</tbody>
</table>

#### Asthma Dataset

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>CITR</td>
<td>74.38</td>
<td>13.87</td>
<td>11.81</td>
<td>7.23</td>
<td>6</td>
</tr>
<tr>
<td>IPTL</td>
<td>71.88</td>
<td>18.58</td>
<td>11.83</td>
<td>7.87</td>
<td>8</td>
</tr>
<tr>
<td>MCA</td>
<td>70.32</td>
<td>19.53</td>
<td>12.01</td>
<td>10.23</td>
<td>-</td>
</tr>
<tr>
<td>OWL (Gaussian)</td>
<td>71.02</td>
<td>25</td>
<td>12.38</td>
<td>16</td>
<td>-</td>
</tr>
<tr>
<td>OWL (Linear)</td>
<td>71.02</td>
<td>25</td>
<td>12.38</td>
<td>16</td>
<td>-</td>
</tr>
<tr>
<td>CITR - No Treat</td>
<td>77.39</td>
<td>14.02</td>
<td>12.87</td>
<td>7.38</td>
<td>7</td>
</tr>
<tr>
<td>CITR - No Assess</td>
<td>78.32</td>
<td>18.28</td>
<td>12.02</td>
<td>8.97</td>
<td>9</td>
</tr>
<tr>
<td>CITR - Outcome</td>
<td>79.37</td>
<td>18.28</td>
<td>12.88</td>
<td>9.21</td>
<td>9</td>
</tr>
</tbody>
</table>

| Human            | 69.37        | -                | 33.39           | -                  | -        |

|                | 68.32        | -                | 12.28           | -                  | -        |
## Zhang et al. Sim Results

<table>
<thead>
<tr>
<th>Scenario</th>
<th>$n$</th>
<th>DL</th>
<th>$Q$-lasso</th>
<th>$Q$-RF</th>
<th>BOWL</th>
<th>SOWL</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>100</td>
<td>6.63 (0.24)</td>
<td>6.55 (0.58)</td>
<td>6.70 (0.05)</td>
<td>6.70 (0.05)</td>
<td>6.70 (0.05)</td>
</tr>
<tr>
<td>I</td>
<td>200</td>
<td>6.73 (0.24)</td>
<td>6.64 (0.33)</td>
<td>6.70 (0.05)</td>
<td>6.70 (0.05)</td>
<td>6.70 (0.05)</td>
</tr>
<tr>
<td>I</td>
<td>400</td>
<td>6.94 (0.16)</td>
<td>6.66 (0.26)</td>
<td>6.70 (0.05)</td>
<td>6.70 (0.05)</td>
<td>6.70 (0.05)</td>
</tr>
<tr>
<td>II</td>
<td>100</td>
<td>3.66 (0.10)</td>
<td>3.68 (0.08)</td>
<td>3.41 (0.17)</td>
<td>3.15 (0.05)</td>
<td>2.77 (0.52)</td>
</tr>
<tr>
<td>II</td>
<td>200</td>
<td>3.71 (0.04)</td>
<td>3.73 (0.04)</td>
<td>3.62 (0.12)</td>
<td>3.22 (0.08)</td>
<td>2.84 (0.33)</td>
</tr>
<tr>
<td>II</td>
<td>400</td>
<td>3.73 (0.03)</td>
<td>3.75 (0.02)</td>
<td>3.71 (0.04)</td>
<td>3.37 (0.14)</td>
<td>2.91 (0.28)</td>
</tr>
<tr>
<td>III</td>
<td>100</td>
<td>14.49 (2.77)</td>
<td>5.42 (4.54)</td>
<td>12.94 (2.07)</td>
<td>10.65 (2.40)</td>
<td>10.27 (2.33)</td>
</tr>
<tr>
<td>III</td>
<td>200</td>
<td>17.42 (1.42)</td>
<td>7.88 (1.63)</td>
<td>15.79 (1.59)</td>
<td>13.09 (2.20)</td>
<td>12.98 (1.88)</td>
</tr>
<tr>
<td>III</td>
<td>400</td>
<td>18.60 (0.71)</td>
<td>8.41 (0.65)</td>
<td>18.02 (0.73)</td>
<td>15.33 (1.56)</td>
<td>16.22 (1.58)</td>
</tr>
<tr>
<td>IV</td>
<td>100</td>
<td>13.38 (3.14)</td>
<td>4.54 (5.17)</td>
<td>11.47 (2.31)</td>
<td>6.72 (1.71)</td>
<td>6.04 (2.18)</td>
</tr>
<tr>
<td>IV</td>
<td>200</td>
<td>17.33 (1.87)</td>
<td>7.69 (2.33)</td>
<td>14.82 (1.75)</td>
<td>8.90 (1.13)</td>
<td>8.34 (1.99)</td>
</tr>
<tr>
<td>IV</td>
<td>400</td>
<td>18.84 (0.70)</td>
<td>8.61 (0.96)</td>
<td>17.04 (1.02)</td>
<td>10.75 (0.68)</td>
<td>9.38 (2.30)</td>
</tr>
<tr>
<td>V</td>
<td>100</td>
<td>23.68 (1.09)</td>
<td>12.97 (3.40)</td>
<td>17.83 (1.63)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>V</td>
<td>200</td>
<td>25.94 (0.51)</td>
<td>13.80 (2.57)</td>
<td>21.60 (1.28)</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>