Doubly Robust Prediction and Evaluation Methods Improve Uplift Modeling for Observational Data

Yuta Saito, Hayato Sakata, and Kazuhide Nakata

yuta_saito@so-netmedia.jp, hayato_sakata@so-netmedia.jp, nakata.k.ac@m.titech.ac.jp

Motivation

• Achieving optimal treatment assignments
  • Medical Treatment, Advertisement, Coupon Distribution

  Should We Treat Them?

Bob or Mary

If we know the optimal treatment of each individual, we could achieve the best possible future (highest survival rates)

Bob or Mary

Problem Setting

• Uplift Modeling tries to find an optimal treatment by analyzing the causal effect using Potential Outcomes

\[
\begin{align*}
Y_{Bob}^{(1)} & \quad \text{Survival Rate with Treatment} \\
Y_{Bob}^{(0)} & \quad \text{Survival Rate without Treatment}
\end{align*}
\]

Goal: Predict the Individual Treatment Effect (ITE)

\[ \tau_{Bob} = Y_{Bob}^{(1)} - Y_{Bob}^{(0)} \]

We have 2 options for gathering training and test data

RCT randomized treatments for data gathering

Pros: Treatments and features are independent
Cons: Cost and time ineffective

Observational historical log data depending on past policies

Pros: Cost and time effective
Cons: Treatment assignments depend on past policies

In this work, we focused on observational data, which is generally available and we can extend the applications

Related Work

• Transformed Outcome (TO) as proxy ITE [Athey+ 2015]

\[
\begin{align*}
Y_{TO}^{i} &= \frac{W_{i}Y_{i}^{obs}}{\epsilon_{i}} - 1 - \frac{W_{i}Y_{i}^{obs}}{1 - \epsilon_{i}} \\
Y_{TO}^{i} \text{ is the observed outcome} \\
W_{i} \in \{0, 1\} \text{ is the treatment assignment indicator} \\
\epsilon_{i} = P(W_{i} = 1 \mid X_{i}) \text{ is the true propensity score} \\
\text{TO is an unbiased estimator for the ITE} \quad \text{[Athey+ 2015]}
\end{align*}
\]

Unbiasedness of the TO is desirable, but...

• True Propensity Score is often missing and TO can be biased with an estimated propensity score

• Variance of TO has yet to be analyzed thus TO can be inaccurate proxy ITE

Proposed Techniques

• Doubly Robust Estimation

Incorporate Potential Outcome Models into TO

\[
\begin{align*}
Y_{i}^{DR} &= \frac{W_{i}Y_{i}^{obs} - \bar{\mu}_{i}^{(0)}}{\epsilon_{i} - 1 - \epsilon_{i}} \\
\bar{\mu}_{i}^{(0)} &\text{ are predicted values of } Y_{i}^{(0)}, Y_{i}^{(k)}
\end{align*}
\]

Bias Analysis

\[
\begin{align*}
\text{Bias}(Y_{TO}^{i} \mid X_{i}) &= |\beta^{(1)}(\tau_{i}^{(1)}) - \beta^{(0)}| \\
\text{Expectations of Potential Outcome} \\
\text{Bias}(Y_{TO}^{i} \mid X_{i}) &= |\beta^{(1)}(\tau_{i}^{(1)}) - \beta^{(0)}| \\
\text{Our Method} \\
\end{align*}
\]

Proposed Proxy: Switch Doubly Robust Outcome

\[
Y_{SDR}^{DR}(y) = \begin{cases} 
\bar{\mu}_{i}^{(1)} - \bar{\mu}_{i}^{(0)} & \text{for Extreme Propensity Score} \\
\text{for } \bar{\mu}_{i} = 1 & \epsilon_{i} < y \\
\bar{\mu}_{i}^{(0)} & \text{or } W_{i} = 0.8 \text{ or } 0.1 < \epsilon_{i} \\
\end{cases}
\]

Unbalanced Propensity Score

Synthetic Experiment

Setup

• Used 8 data generating processes from [Powers+ 2017]

• For prediction methods:
  • Compared TMA, TRM, SDRM (\(y = 0.0, 0.3, 0.5\)) by ITE prediction performance

• For evaluation metrics:
  • Compared \(\mu\text{-risk, } x\text{-risk, TO-MSE, } SDR\text{-MSE} (y = 0.0, 0.3, 0.5)\) by model selection performance

Results

Prediction Evaluation

• Our prediction method (SDRM) demonstrated the best prediction accuracies in all scenarios

• 0.5 is the optimal value for hyper-parameter \(y\)

• The effect of varying \(y\) is relatively small but a positive value is better than zero

Real-World Experiment

Setup

• Right Heart Catheterization (RHC) data
  • well-known public dataset
  • 5,735 critically ill patients
  • Average Treatment Effect of RHC was found to be negative

Results

• Ours found 20% of positively affected patients

Uplift Curve

x-axis: Ratio of Treated Individuals

y-axis: Difference of Survival rates between the treated and the controlled