Identify Qualitative Interaction Through Value of Information

Peng Zhang

Institute for Social Research
University of Michigan

Joint work with James Robins and Susan Murphy
Goal

- Identifying variables that are important for adapting or personalizing treatment
- Interested in variables that have qualitative interaction with the treatment
- Variables that qualitatively interact with treatment
  - Inform us about the magnitude of treatment effect
  - Differentiate between patients who should be offered different treatments

- Traditional variable selection techniques focus on variable selection for the prediction of the response in a supervised learning setting.
- We propose to identify such variables through value of information
Goal

- Identifying variables that are important for adapting or personalizing treatment
- Interested in variables that have qualitative interaction with the treatment
- Variables that qualitatively interact with treatment
  - Inform us about the magnitude of treatment effect
  - Differentiate between patients who should be offered different treatments
- Traditional variable selection techniques focus on variable selection for the prediction of the response in a supervised learning setting.
- We propose to identify such variables through value of information
Notations

- Observe i.i.d. $O = (Y, A, X)$
- $Y$ is the clinical outcome (the larger the better).
- $X = (X_1, X_2, \ldots, X_p)$ is the $p$-dimensional covariates.
- $A$ is the treatment indicator with known randomization probabilities.
- We simply assume that $P(A = 1 \mid X) = P(A = -1 \mid X) = \frac{1}{2}$.
Value of Information

Marginal Optimal Decision

\[ A^{opt} = \arg \max_a E(Y | A = a) \]

Optimal Reward

\[ E(Y | A = A^{opt}) = \max \{ E(Y | A = 1), E(Y | A = -1) \} \]
Value of Information

Marginal Optimal Decision

$$A^{opt} = \arg \max_a E(Y \mid A = a)$$

Optimal Reward

$$E(Y \mid A = A^{opt}) = \max \{ E(Y \mid A = 1), E(Y \mid A = -1) \}$$
Value of Information (Cont’d)

- Very few of variables are likely to be useful for deciding which treatment to provide to which patient.
- We consider the linear decision rule based on a single covariate.

**Optimal Reward**

$$\sup_{\beta_0, \beta_1} E (E (Y \mid A = \text{sgn}(\beta_0 + \beta_1 X_1), X_1))$$
Value of Information (Cont’d)

- Very few of variables are likely to be useful for deciding which treatment to provide to which patient.
- We consider the linear decision rule based on a single covariate.

Optimal Reward

$$\sup_{\beta_0, \beta_1} E (E (Y \mid A = \text{sgn} (\beta_0 + \beta_1 X_1), X_1))$$
Value of Information (Cont’d)

The value of this additional information $X_1$ for making the decision of treatment is

\[
\theta_1 \triangleq \sup_{\beta_0, \beta_1} E\left( E(Y \mid A = \text{sgn}(\beta_0 + \beta_1 X_1), X_1) \right) - \max\left\{ E(Y \mid A = 1), E(Y \mid A = -1) \right\}
\]

- The bigger VoI is, the more useful the additional information is.
- One might agree that is important for making decision when is greater than the clinical significance of the clinical outcome.
- We will identify $X_1$ whenever $\theta_1 > 0$, and define it as a QI.
- We want to test $H_0 : \theta_1 = 0$ vs $H_1 : \theta_1 > 0$. 
\[ \theta_1 = \min \left\{ \sup_\beta EU_{11}(\beta), \sup_\beta EU_{12}(\beta) \right\} \]

where

\[ U_{11}(\beta) = -2YAI(\beta_0 + \beta_1 X_1 < 0) \]
\[ U_{12}(\beta) = 2YAI(\beta_0 + \beta_1 X_1 \geq 0) \]

Notice that \( U_{11}(\beta) = -U_{12}(\beta) \).
Inference on Non-regular Parameter

• $\mu_{11}(\beta) = E(U_{11}(\beta))$ and $\mu_{12}(\beta) = E(U_{12}(\beta))$ are regular parameters.

• The operator $(f(\beta), g(\beta)) \rightarrow \min \{\sup_\beta f, \sup_\beta g\}$ is not differentiable. Hence, functional delta method does not apply in this setting.

• $\theta_1$ is a non-regular parameter.

• We will construct confidence interval, possibly conservative, for $\theta_1$ at the design level.
Inference on Non-regular Parameter

- \( \mu_{11}(\beta) = E(U_{11}(\beta)) \) and \( \mu_{12}(\beta) = E(U_{12}(\beta)) \) are regular parameters.
- The operator \((f(\beta), g(\beta)) \rightarrow \min \{ \sup_{\beta} f, \sup_{\beta} g \} \) is not differentiable. Hence, functional delta method does not apply in this setting.

- \( \theta_1 \) is a non-regular parameter.
- We will construct confidence interval, possibly conservative, for \( \theta_1 \) at the design level.
Projection Interval

If we can construct $1 - \alpha$ level confidence interval $Cl_\omega$ for parameter $\omega$, for any function $h$ and $\varphi = h(\omega)$, we can obtain the projection interval $Cl_\varphi$ which is the image of $h$ on $Cl_\omega$. $Cl_\varphi$ will be an at-least $1 - \alpha$ level confidence interval for $\varphi$.

- $\omega$: $\mu_{11}(\beta)$ and $\mu_{12}(\beta)$
- $h(f(\beta), g(\beta)) \rightarrow \min \{ \sup_\beta f, \sup_\beta g \}$
- $\varphi$: $\theta_1$
Projection Interval

If we can construct $1 - \alpha$ level confidence interval $CI_\omega$ for parameter $\omega$, for any function $h$ and $\varphi = h(\omega)$, we can obtain the projection interval $CI_\varphi$ which is the image of $h$ on $CI_\omega$. $CI_\varphi$ will be an at-least $1 - \alpha$ level confidence interval for $\varphi$.

- $\omega$: $\mu_{11}(\beta)$ and $\mu_{12}(\beta)$
- $h(f(\beta), g(\beta)) \rightarrow \min \{\sup_\beta f, \sup_\beta g\}$
- $\varphi$: $\theta_1$
Adaption

Consider confidence interval for $\mu_{11}(\beta)$ and $\mu_{12}(\beta)$ of the following form:

$$P(B_{11}(\beta) < \mu_{11}(\beta), B_{12}(\beta) < \mu_{12}(\beta) \text{ for all } \beta) = 1 - \alpha$$

Uniform confidence interval

$$P(\sup \sqrt{n}(P_n - P) U_{11}(\beta) < \lambda, \sup \sqrt{n}(P_n - P) U_{12}(\beta) = 1 - \alpha$$
Adaption (Cont’d)

Adaptive confidence interval

\[ P(\sup \sqrt{n}(P_n - P) U_{11}(\beta) < \lambda \, w_1(\beta), \sup \sqrt{n}(P_n - P) U_{12}(\beta) < \lambda \, w_2(\beta)) = 1 - \alpha \]

where \( w_1(\beta) = \exp \left\{ -c \max (|U_{11}(\beta)|, |U_{12}(-\beta)|) \right\} \) and
\( w_2(\beta) = \exp \left\{ -c \max (|U_{12}(\beta)|, |U_{11}(-\beta)|) \right\} \)
Summary

- We provide a general framework to identify qualitative interactions through value of information.
- This extends original definition from the seminal paper by Gail and Simon to continuous covariates.
- We use adaptive projection interval to deal with inference on non-regular parameters.
Future Work

- Use Bayesian to improve the power of the test
- Extend it to the multi-stage problem