On Evaluation Of Personalized Intervention Algorithms

Drafted by Haoda Fu, Ph.D.

Eli Lilly and Company

November 11, 2018

Advanced Analytics and Data Sciences (AADS)
Eli Lilly and Company

Outline

1. Examples of personalized intervention
2. Optimal rules
3. Training data
4. Testing data
5. Evaluation process and criteria
6. Example

Where is the value from

Context: Digital technology will enable us to collect more individual patient data.

Where is the value from

Context: Digital technology will enable us to collect more individual patient data.

Decision: The purpose of collecting these data is to generate actionable insights.

Where is the value from

Context: Digital technology will enable us to collect more individual patient data.

Decision: The purpose of collecting these data is to generate actionable insights.

Reward: The goal of these actionable insights is to maximize individual patient's outcomes.

Learning - from data to knowledge

Human beings, mammals learn from experiences. Experiences are data containing Context, Decision, and Reward .

What we have: Collect *Context*, *Decision*, *Reward* from previous cases.

What we want to do: Develop *algorithms* to figure out optimal decisions on existing data.

How we apply it: Apply the algorithms to recommend the optimal decision for a new context.

Learning - from data to knowledge

Human beings, mammals learn from experiences. Experiences are data containing Context, Decision, and Reward .

What we have: Collect *Context*, *Decision*, *Reward* from previous cases.

What we want to do: Develop *algorithms* to figure out optimal decisions on existing data.

How we apply it: Apply the algorithms to recommend the optimal decision for a new context.

This topic is important because it is *the* framework to make optimal decision based on data.

Illustration Data

Table: An illustration dataset

ID	Y	A	X_1	X_2	X_3	$\cdots$
1	1.5	1	F	26	7.8	$\cdots$
2	1.2	2	M	28	8.2	$\cdots$
3	2.3	3	M	31	8.9	$\cdots$
4	0.9	2	F	35	9.4	$\cdots$
5	1.7	1	M	22	7.3	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

Research Question

Based on these data, how can we treat a new patient?
In other words, how can we learn a treatment assignment rule that, if followed by the entire population of certain patients, would lead to the best outcome on average?

Other Examples: Car Purchase

Table: My Friends' Rating of Their First Cars

ID	Satisfaction	Car Type	Gender	Age	Mileage per Day	⋯
1	90%	Focus	F	26	7.8	⋯
2	85%	Corolla	M	28	8.2	⋯
3	70%	Civic	M	31	8.9	⋯
4	75%	Corolla	F	35	9.4	⋯
5	60%	Civic	M	22	7.3	⋯
⋮	⋮	⋮	⋮	⋮	⋮	⋱

Learning from these data, what car should I purchase?

Lilly

Other Examples: Connected Care Device

Table: Sending Out a Reminder at Right Time for Right Patients

ID	Cost	Send Reminder	FBG	3 Hypo	SU	$\cdots$
1	\$875	0	159	Y	Y	$\cdots$
2	\$475	0	170	Y	N	$\cdots$
3	\$150	1	160	N	N	$\cdots$
4	\$375	1	182	Y	Y	$\cdots$
5	\$525	1	110	N	Y	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

Learning from these data, how can we develop a smart reminder to recommend patients to see their doctor within the next 3 weeks?

Other Examples: Choice of Digital Biomarkers

Table: Choose Right Digital Biomarker for Alzheimer's Disease

ID	Accuracy	Digital Biomarker	State	Age	Gender	$\cdots$
1	70%	App No.1	Mild	63	F	$\cdots$
2	83%	App No.2	Moderate	72	F	$\cdots$
3	77%	App No.1	Mild	65	M	$\cdots$
4	62%	App No.3	Severe	86	M	$\cdots$
5	53%	App No.2	Moderate	77	F	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

Learning from these data, which is the most accurate digital biomarker that we need to choose for a new patient based on this subject's characteristics? If we can only choose one digital biomarker for patients with mild Alzheimer's Disease which one we need to utilize?

Other Examples: Business Investment

Table: Previous Commercial Investments and Returns

Case ID	Return	Type	Month	Location	Share of Market	$\cdots$
1	1.2	TV	Jan	MW	12.5	$\cdots$
2	0.9	Radio	Oct	NE	18.2	$\cdots$
3	1.4	Web	Nov	WE	12.9	$\cdots$
4	1.3	Web	Dec	MW	10.4	$\cdots$
5	1.2	Radio	Feb	SE	11.3	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

Learning from these data, what is our best way to invest in New England area if our product has 12% market share in this March?

The following problems have a common theme

- Individualized treatment recommendation
- Reminder system
- Clinical decision system
- Recommender system
- Intelligent assistant
- Multichannel marketing
- *Subgroup identification*

Outline

1. Examples of personalized intervention
2. Optimal rules
3. Training data
4. Testing data
5. Evaluation process and criteria
6. Example

Notation

- $Y \in \mathbb{R}$: a response.
- $X \in \mathbb{R}^p$: a vector of covariates.
- $A \in \mathbb{Z}$: an action.
- $\mathcal{X}$: population space, i.e. $\mathcal{X} = \{X | X \in \mathbb{R}^p\} \subseteq \mathbb{R}^p$.
- $\mathcal{A}$: an action space, i.e. $\mathcal{A} = \{A | A \in \mathbb{Z}\} \subseteq \mathbb{Z}$.
- Lower case of a is an realization of A.
- E: expectation.

Optimal rule and the Fisher's Consistency

Optimal Rule

$$\mathcal{D}^*(X) = \underset{a \in \mathcal{A}}{\operatorname{argmax}}\, E\{Y|X, A = a\}$$

Optimal rule and the Fisher's Consistency

Optimal Rule

$$\mathcal{D}^*(X) = \operatorname*{argmax}_{a \in \mathcal{A}} E\{Y|X, A = a\}$$

The Fisher's Consistency

$$\widehat{\mathcal{D}}(X) \xrightarrow{\mathcal{P}} \mathcal{D}^*(X), \quad \forall X \in \mathcal{X}.$$

Data generation models

Generalized Linear Model:

$$\ell\{E(Y|X)\} = \beta_0 + g(X) + t(A) + d(X, A),$$

where $\ell(\cdot)$ is a monotone link function.

Transformed Response Model:

$$\tau(Y) = \beta_0 + g(X) + t(A) + d(X, A) + \epsilon,$$

where τ is a monotone transformation function, and $\epsilon \sim (0, \sigma^2)$ which can be nonparametric.

Intensity Function Model:

$$\lambda_i(t) = \lambda_0(t)\gamma_i \exp\{g(X_i) + t(A) + d(X_i, A)\},$$

where $\lambda_0(t)$ is a baseline hazard or intensity function, and γ_i is a frailty term, e.g. $\gamma_i \sim \text{Gamma}(1, \sigma^2)$.

Outline

1 Examples of personalized intervention

2 Optimal rules

3 Training data

4 Testing data

5 Evaluation process and criteria

6 Example

Training data

Table: An illustration training dataset

ID	Y	A	X_1	X_2	X_3	$\cdots$
1	1.5	1	0	26	7.8	$\cdots$
2	1.2	2	1	28	8.2	$\cdots$
3	0.3	3	1	31	8.9	$\cdots$
4	0.9	2	0	35	9.4	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$
n	1.6	2	0	29	8.1	$\cdots$

Research Question

Based on these data, how can we treat a new patient?
In other words, how can we learn a treatment assignment rule that, if followed by the entire population of certain patients, would lead to the best outcome on average?

Outline

1. Examples of personalized intervention
2. Optimal rules
3. Training data
4. Testing data
5. Evaluation process and criteria
6. Example

Testing data set: Context

Table: *An illustration testing dataset: context data (testing)*

ID	X_1	X_2	X_3	$\cdots$
1	0	26	7.8	$\cdots$
2	1	28	8.2	$\cdots$
3	1	31	8.9	$\cdots$
4	0	35	9.4	$\cdots$
5	1	22	7.3	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$
N	0	29	8.1	$\cdots$

Testing data set: Potential outcomes

Table: *An illustration testing dataset: potential outcome data(testing).* $Y(a)$ is the potential outcome taking action a, A is the observed treatment assignment, and A^o is the theoretical optimal treatment assignment.

ID	$Y(1)$	$Y(2)$	$\cdots$	$Y(k)$	A	A^o
1	1.2	1.5	$\cdots$	1.3	3	2
2	1.3	1.1	$\cdots$	1.4	2	3
3	0.9	0.8	$\cdots$	1.7	1	3
4	1.8	1.6	$\cdots$	1.2	1	1
5	1.4	1.4	$\cdots$	1.5	2	2
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$
N	1.7	1.4	$\cdots$	1.1	3	1

Evaluation Process

We (simulator) have access to all potential outcomes for each patient (in real data only one outcome is observed). Suppose, we would like to evaluate a $\mathcal{D}(\cdot) : \mathcal{X} \mapsto \mathcal{A}$.

Step 1: We send training data to the modelers, and they train a decision rule $\widehat{\mathcal{D}}(\cdot)$.

Step 2: We send the context data (testing) to the modelers, and they apply $\widehat{\mathcal{D}}(\cdot)$ on the context data (testing), and they send us a vector $\widehat{A}$ with action for each subject.

Step 3: We use their $\widehat{A}$ and potential outcome (testing) data to calculate scores based on different evaluation criteria.

Evaluation criteria

Primary criteria: The average benefit $N^{-1}\sum_{i=1}^{N}\sum_{a=1}^{k} Y_i(a) I\{a=\widehat{A}_i\}$.

Secondary criteria 1: Proportion of misclassification $N^{-1}\sum_{i=1}^{N} I\{A_i^o \neq \widehat{A}_i\}$.

Secondary criteria 2: Average of proportion of misclassification

$$k^{-1}\sum_{a=1}^{k}\left\{\sum_{i=1}^{N} I(A_i^o=a)\right\}^{-1}\sum_{i=1}^{N} I\{A_i^o \neq \widehat{A}_i, A_i^o=a\}.$$

Outline

1. Examples of personalized intervention
2. Optimal rules
3. Training data
4. Testing data
5. Evaluation process and criteria
6. Example

Example

Table: *An illustration of using the potential outcome data (testing) to evaluate solution.*

ID	$Y(1)$	$Y(2)$	A	A^o	$\widehat{A}$
1	1.2	1.5	1	2	2
2	1.3	1.1	2	1	1
3	0.9	0.8	1	2	1
4	1.8	1.6	2	1	2
5	1.4	1.4	2	2	2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
N	1.7	1.4	2	1	1

The observed benefit is $V = \frac{1}{N}\sum(1.2 + 1.1 + 0.9 + 1.6 + 1.4 + \cdots + 1.4)$, and theoretical optimal value is $V^o = \frac{1}{N}\sum(1.5 + 1.3 + 0.8 + 1.8 + 1.4 + \cdots + 1.7)$, and the estimated value is $\widehat{V} = \frac{1}{N}\sum(1.5 + 1.3 + 0.9 + 1.6 + 1.4 + \cdots + 1.7)$.