The Role of Active Learning in Sequential Decision Making

Daniel Lizotte
University of Waterloo

Plan

- Discuss "Active Learning" background
- Formalize "**Active Action Choice**" framework
- Propose an algorithm for AAC, give Bad News and Good News

“Active Learning”

- Optimal Experimental Design
- Focuses on predictive performance
- Many different settings
- Terminology has not converged

"Active Learning"

x_1	x_2	x_3	...	x_p	y
0	1	0	...	0	
0	1	1	...	1	
1	0	1	...	1	
1	1	1	...	1	
0	1	0	...	0	
0	0	0	...	0	
1	1	1	...	1	
...	...	...	...	...	...

“Active Learning”

x_1	x_2	x_3	...	x_p	y
0	1	0	...	0	
0	1	1	...	1	1
1	0	1	...	1	
1	1	1	...	1	
0	1	0	...	0	
0	0	0	...	0	
1	1	1	...	1	
...	...	...	...	...	...

"Active Learning"

x_1	x_2	x_3	...	x_p	y
0	1	0	...	0	0
0	1	1	...	1	1
1	0	1	...	1	
1	1	1	...	1	
0	1	0	...	0	
0	0	0	...	0	
1	1	1	...	1	
...	...	...	...	...	...

“Active Learning”

x_1	x_2	x_3	...	x_p	y
0	1	0	...	0	0
0	1	1	...	1	1
1	0	1	...	1	
1	1	1	...	1	
0	1	0	...	0	
0	0	0	...	0	1
1	1	1	...	1	
...	...	...	...	...	...

“Active Learning”

x_1	x_2	x_3	...	x_p	y
0	1	0	...	0	0
0	1	1	...	1	1
1	0	1	...	1	
1	1	1	...	1	1
0	1	0	...	0	
0	0	0	...	0	1
1	1	1	...	1	
...	...	...	...	...	...

“Active Learning”

x_1	x_2	x_3	...	x_p	y
0	1	0	...	0	0
0	1	1	...	1	1
1	0	1	...	1	0
1	1	1	...	1	1
0	1	0	...	0	1
0	0	0	...	0	1
1	1	1	...	1	0
...	...	...	...	...	...

“Budgeted Learning”

x_1	x_2	x_3	...	x_p	y
			...		0
			...		1
			...		0
			...		1
			...		1
			...		1
			...		0
...	...	...	...	...	...

“Budgeted Learning”

x_1	x_2	x_3	...	x_p	y
0			...		0
			...		1
			...		0
			...		1
			...		1
			...		1
			...		0
...	...	...	...	...	...

“Budgeted Learning”

x_1	x_2	x_3	...	x_p	y
0		0	...		0
			...		1
			...		0
			...		1
			...		1
			...		1
			...		0
...	...	...	...	...	...

“Budgeted Learning”

x_1	x_2	x_3	...	x_p	y
0		0	...		0
			...		1
			...		0
		1	...		1
			...		1
			...		1
			...		0
...	...	...	...	...	...

“Budgeted Learning”

x_1	x_2	x_3	...	x_p	y
0		0	...		0
			...		1
			...		0
		1	...	1	1
			...		1
			...		1
			...		0
...	...	...	...	...	...

"Budgeted Learning"

x_1	x_2	x_3	...	x_p	y
0		0	...		0
		1	...		1
			...		0
		1	...	1	1
			...		1
			...		1
			...		0
...	...	...	...	...	...

“Budgeted Learning”

x_1	x_2	x_3	...	x_p	y
0		0	...		0
		1	...	1	1
			...		0
		1	...	1	1
			...		1
			...		1
			...		0
...	...	...	...	...	...

“Budgeted Learning”

x_1	x_2	x_3	...	x_p	y
0		0	...		0
		1	...	1	1
			...		0
		1	...	1	1
			...		1
0			...		1
			...		0
...	...	...	...	...	...

“Budgeted Learning”

x_1	x_2	x_3	...	x_p	y
0		0	...		0
		1	...	1	1
			...		0
		1	...	1	1
			...		1
0		0	...		1
			...		0
...	...	...	...	...	...

“Budgeted Learning”

x_1	x_2	x_3	...	x_p	y
0		0	...	0	0
		1	...	1	1
			...		0
		1	...	1	1
			...		1
0		0	...		1
			...		0
...	...	...	...	...	...

“Budgeted Learning”

x_1	x_2	x_3	...	x_p	y
0		0	...	0	0
	1	1	...	1	1
			...		0
		1	...	1	1
			...		1
0		0	...		1
			...		0
...	...	...	...	...	...

“Budgeted Learning”

x_1	x_2	x_3	...	x_p	y
0		0	...	0	0
	1	1	...	1	1
			...		0
		1	...	1	1
			...		1
0		0	...	0	1
			...		0
...	...	...	...	...	...

“Budgeted Learning”

x_1	x_2	x_3	...	x_p	y
0		0	...	0	0
	1	1	...	1	1
	0		...		0
		1	...	1	1
			...		1
0		0	...	0	1
			...		0
...	...	...	...	...	...

“Budgeted Learning”

x_1	x_2	x_3	...	x_p	y
0		0	...	0	0
	1	1	...	1	1
	0	1	...		0
		1	...	1	1
			...		1
0		0	...	0	1
			...		0
...	...	...	...	...	...

“Budgeted Learning”

x_1	x_2	x_3	...	x_p	y
0		0	...	0	0
	1	1	...	1	1
	0	1	...		0
		1	...	1	1
		0	...		1
0		0	...	0	1
			...		0
...	...	...	...	...	...

"Budgeted Learning"

x_1	x_2	x_3	...	x_p	y
0		0	...	0	0
	1	1	...	1	1
	0	1	...		0
		1	...	1	1
		0	...		1
0	0	0	...	0	1
			...		0
...	...	...	...	...	...

“Budgeted Learning”

x_1	x_2	x_3	...	x_p	y
0		0	...	0	0
	1	1	...	1	1
	0	1	...		0
	1	1	...	1	1
		0	...		1
0	0	0	...	0	1
			...		0
...	...	...	...	...	...

"Budgeted Learning"

x_1	x_2	x_3	...	x_p	y
0		0	...	0	0
	1	1	...	1	1
	0	1	...		0
	1	1	...	1	1
	1	0	...		1
0	0	0	...	0	1
			...		0
...	...	...	...	...	...

"Budgeted Learning"

x_1	x_2	x_3	...	x_p	y
0	1	0	...	0	0
0	1	1	...	1	1
1	0	1	...	1	0
1	1	1	...	1	1
0	1	0	...	0	1
0	0	0	...	0	1
1	1	1	...	1	0
...	...	...	...	...	...

"Active Classification"
"Active Diagnosis"

x_1	x_2	x_3	...	x_p	y
			...		"?"

“Active Classification”
“Active Diagnosis”

x_1	x_2	x_3	…	x_p	y
0			…		“?”

“Active Classification”
“Active Diagnosis”

x_1	x_2	x_3	...	x_p	y
0			...	1	“?”

“Active Classification”
“Active Diagnosis”

x_1	x_2	x_3	…	x_p	y
0		1	…	1	“?”

“Active Classification”
“Active Diagnosis”

x_1	x_2	x_3	...	x_p	y
0	0	1	...	1	“1”

Action Choice in a DTR

- We define $Q(\boldsymbol{x},a)$ to be the expected reward achieved by taking action a in state $\boldsymbol{x} = (x_1, x_2, ..., x_p)$ and following with the optimal policy
- Best action in $\boldsymbol{x}$ is $\pi(a) = \text{argmax}_a\ Q(\boldsymbol{x},a)$
- Assumes $\boldsymbol{x}$ is completely observed

Active Action Choice

x_1	x_2	x_3	...	x_p	a
0	0	1	...	1	"1"

Available at some cost

Txt Action

- Why?
  - Set a budget, still make good decisions
  - Set a bar (regret), be as cost-effective as possible

Necessary Tools

x_1	x_2	x_3	...	x_p	a
0	0	1	...	1	"1"

Available at some cost

Txt Action

Necessary Tools

1. Mechanism for choosing an action when covariates are “missing”

x_1	x_2	x_3	...	x_p	a
0	0	1	...	1	“1”

Available at some cost

Txt Action

Necessary Tools

1. Mechanism for choosing an action when covariates are "missing"
2. Mechanism for deciding which covariate to purchase next

x_1	x_2	x_3	...	x_p	a
0	0	1	...	1	"1"

Available at some cost

Txt Action

Necessary Tools

1. **Mechanism for choosing an action when covariates are "missing"**
2. Mechanism for deciding which covariate to purchase next

x_1	x_2	x_3	...	x_p	a
0	0	1	...	1	"1"

Available at some cost

Txt Action

Action Choice from Missing Covariates

Action Choice from Missing Covariates

- Assumption 1:
  Access to Q($\boldsymbol{x}$,a)

Action Choice from Missing Covariates

- Assumption 1:
  Access to Q($\boldsymbol{x}$,a)
- This is about *deploying* a DTR, not *estimating* a DTR

Action Choice from Missing Covariates

- Assumption 1:
  Access to Q($\boldsymbol{x}$,a)
- This is about *deploying* a DTR, not *estimating* a DTR
- Someone else did the heavy lifting

Action Choice from Missing Covariates

Action Choice from Missing Covariates

- Assumption 2:
  Access to $P(X_1, X_2, \ldots, X_p)$

Action Choice from Missing Covariates

- Assumption 2:
  Access to $P(X_1, X_2, \ldots, X_p)$
- Egregious? Maybe.

Action Choice from Missing Covariates

- Assumption 2:
  Access to $P(X_1,X_2,\ldots,X_p)$
- Egregious? Maybe.
  - No parametric assumptions

Action Choice from Missing Covariates

- Assumption 2:
  Access to $P(X_1,X_2,\ldots,X_p)$
- Egregious? Maybe.
  - No parametric assumptions
  - P represents population "at large"; could estimate from other data

Action Choice from Missing Covariates

- Assumption 2:
  Access to $P(X_1, X_2, \ldots, X_p)$
- Egregious? Maybe.
  - No parametric assumptions
  - *P* represents population "at large"; could estimate from other data
  - Interesting problems down that road

Action Choice from Missing Covariates

- Maximize expected reward, expectation taken over the missing covariates

$$\arg\max_a \mathbb{E}_{\mathbf{X}_{\mathrm{m}}|\mathbf{x}_{\mathrm{o}}}\left[Q((\mathbf{X}_{\mathrm{m}}, \mathbf{x}_{\mathrm{o}}), a)\right]$$

- Write $V(\mathbf{x}_{\mathrm{O}})$ for the value of the above action knowing $\mathbf{x}_{\mathrm{O}}$

Necessary Tools

1. Mechanism for choosing an action when covariates are "missing"
2. **Mechanism for deciding which covariate to purchase next**

x_1	x_2	x_3	…	x_p	a
0	0	1	…	1	"1"

Available at some cost

Txt Action

Which covariate should we purchase next?

- What is the “value” we expect if we decide to pay to reveal x_r?

Which covariate should we purchase next?

- What is the "value" we expect if we decide to pay to reveal x_r?

$$\mathbb{E}_{X_r|\mathbf{x}_o}\left[\max_a \mathbb{E}_{\mathbf{X}_{\mathrm{m}'}|\mathbf{x}_{\mathrm{o}},x_r}\left[Q((\mathbf{X}_{\mathrm{m}'},\mathbf{x}_{\mathrm{o}},x_r),a)\right]\right]$$

Which covariate should we purchase next?

- What is the "value" we expect if we decide to pay to reveal x_r?

$$\mathbb{E}_{X_r|\mathbf{x}_o}\left[\max_a \mathbb{E}_{\mathbf{X}_{\mathrm{m}'}|\mathbf{x}_{\mathrm{o}},x_r}\left[Q((\mathbf{X}_{\mathrm{m}'},\mathbf{x}_{\mathrm{o}},x_r),a)\right]\right]$$

- Write $\mathbb{E}_{X_r|\mathbf{x}_{\mathrm{o}}}\left[V(\mathbf{x}_{\mathrm{o}} \cup X_r)\right]$
- Note: $\mathbb{E}_{X_r|\mathbf{x}_{\mathrm{o}}}\left[V(\mathbf{x}_{\mathrm{o}} \cup X_r)\right] - V(\mathbf{x}_{\mathrm{o}}) \geq 0$

Which covariate should we purchase next?

$$\mathbb{E}_{X_r|\mathbf{x}_\mathrm{o}}\left[V(\mathbf{x}_\mathrm{o} \cup X_r)\right] = \mathbb{E}_{X_r|\mathbf{x}_\mathrm{o}}\left[\max_a \mathbb{E}_{\mathbf{X}_{\mathrm{m}'}|\mathbf{x}_\mathrm{o},x_r}\left[Q((\mathbf{X}_{\mathrm{m}'}, \mathbf{x}_\mathrm{o}, x_r), a)\right]\right]$$

Which covariate should we purchase next?

$$\mathbb{E}_{X_r|\mathbf{x}_\mathrm{o}}\left[V(\mathbf{x}_\mathrm{o} \cup X_r)\right] = \mathbb{E}_{X_r|\mathbf{x}_\mathrm{o}}\left[\max_a \mathbb{E}_{\mathbf{X}_{\mathrm{m}'}|\mathbf{x}_\mathrm{o},x_r}\left[Q((\mathbf{X}_{\mathrm{m}'}, \mathbf{x}_\mathrm{o}, x_r), a)\right]\right]$$

- If we will only reveal *one* more covariate, the optimal choice is to reveal $\arg\max_r \mathbb{E}_{X_r|\mathbf{x}_\mathrm{o}}\left[V(\mathbf{x}_\mathrm{o} \cup X_r)\right]$

Which covariate should we purchase next?

$$\mathbb{E}_{X_r|\mathbf{x}_\mathrm{o}}\left[V(\mathbf{x}_\mathrm{o} \cup X_r)\right] = \mathbb{E}_{X_r|\mathbf{x}_\mathrm{o}}\left[\max_a \mathbb{E}_{\mathbf{X}_{\mathrm{m}'}|\mathbf{x}_\mathrm{o},x_r}\left[Q((\mathbf{X}_{\mathrm{m}'}, \mathbf{x}_\mathrm{o}, x_r), a)\right]\right]$$

- If we will only reveal *one* more covariate, the optimal choice is to reveal $\arg\max_r \mathbb{E}_{X_r|\mathbf{x}_\mathrm{o}}\left[V(\mathbf{x}_\mathrm{o} \cup X_r)\right]$
- What if we plan to reveal several more?

Purchasing Policies

- Purchase x_1
  - If $x_1 > 0.743$, purchase x_2
    - If $x_1 * x_2 < 0.4$, purchase x_4
    - ...
  - Else purchase x_3
    - If $x_1 + x_3 > 3.223$, purchase x_2
    - ...

Purchasing Policies

- Purchase x_1
  - If $x_1 > 0.743$, purchase x_2
    - If $x_1 * x_2 < 0.4$, purchase x_4
    - ...
  - Else purchase x_3
    - If $x_1 + x_3 > 3.223$, purchase x_2
    - ...
- **Optimal purchasing is a DTR**

“Turtles all the way down?”

- You: “You’re really going to make us estimate another DTR to deploy the one we already have?”
- Me: “Maybe…”

- This DTR has a lot of structure; can we exploit it? *Might the optimal pruchasing policy have simple structure?*

How good is the greedy policy?

- What if we repeatedly reveal the x_r for which $\mathbb{E}_{X_r|\mathbf{x}_o}[V(\mathbf{x}_o \cup X_r)]$ is maximized?
- It is known that this policy is approximately optimal if the objective is *adaptive submodular*.

Adaptive Submodularity

Adaptive Submodularity

- "Expected benefit from revealing x_r <u>now</u> is at least as high as revealing it <u>in the future</u>."

Adaptive Submodularity

- "Expected benefit from revealing x_r now is at least as high as revealing it in the future."

- If $\mathbf{x}_\mathrm{o} \subseteq \mathbf{x}_{\mathrm{o}'}$, then

$$\mathbb{E}_{X_r|\mathbf{x}_\mathrm{o}}\left[V(\mathbf{x}_\mathrm{o} \cup X_r)\right] - V(\mathbf{x}_\mathrm{o}) \geq \mathbb{E}_{X_r|\mathbf{x}_{\mathrm{o}'}}\left[V(\mathbf{x}_{\mathrm{o}'} \cup X_r)\right] - V(\mathbf{x}_{\mathrm{o}'})$$

Adaptive Submodularity

- "Expected benefit from revealing x_r now is at least as high as revealing it in the future."
- If $\mathbf{x}_{\mathrm{o}} \subseteq \mathbf{x}_{\mathrm{o}'}$, then

$$\mathbb{E}_{X_r|\mathbf{x}_{\mathrm{o}}}[V(\mathbf{x}_{\mathrm{o}} \cup X_r)] - V(\mathbf{x}_{\mathrm{o}}) \geq \mathbb{E}_{X_r|\mathbf{x}_{\mathrm{o}'}}[V(\mathbf{x}_{\mathrm{o}'} \cup X_r)] - V(\mathbf{x}_{\mathrm{o}'})$$

- Limits the "interaction" among feature-reveals

Adaptive Submodularity

- "Expected benefit from revealing x_r now is at least as high as revealing it in the future."

- If $\mathbf{x}_{\mathrm{o}} \subseteq \mathbf{x}_{\mathrm{o}'}$, then

$$\mathbb{E}_{X_r|\mathbf{x}_{\mathrm{o}}}[V(\mathbf{x}_{\mathrm{o}} \cup X_r)] - V(\mathbf{x}_{\mathrm{o}}) \geq \mathbb{E}_{X_r|\mathbf{x}_{\mathrm{o}'}}[V(\mathbf{x}_{\mathrm{o}'} \cup X_r)] - V(\mathbf{x}_{\mathrm{o}'})$$

- Limits the "interaction" among feature-reveals
- **Greedy optimization of adaptive submodular functions is (1 - 1/e) ≈ 0.632 of optimal**

Results for Active Action Choice

Results for Active Action Choice

- Is the general objective adaptive submodular?

Results for Active Action Choice

- Is the general objective adaptive submodular?
  - No.

Results for Active Action Choice

- Is the general objective adaptive submodular?
  - No.
- What if Q is linear in $\boldsymbol{x}$?

Results for Active Action Choice

- Is the general objective adaptive submodular?
  - No.
- What if Q is linear in $\boldsymbol{x}$?
  - No.

Results for Active Action Choice

- Is the general objective adaptive submodular?
  - No.
- What if Q is linear in $\boldsymbol{x}$?
  - No.
- Okay what if Q is linear in $\boldsymbol{x}$ <u>and</u> the X_i are all independent?

Results for Active Action Choice

- Is the general objective adaptive submodular?
  - No.
- What if Q is linear in $\boldsymbol{x}$?
  - No.
- Okay what if Q is linear in $\boldsymbol{x}$ <u>and</u> the X_i are all independent?
  - No.

How bad can it be?

How bad can it be?

- Pretty bad.

How bad can it be?

- Pretty bad.

$$P(X_1, X_2, X_3) \sim \mathcal{N}(\mu, \Sigma)$$

$$\mu = (0, 0, 0)^{\mathsf{T}}$$

$$\Sigma = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -0.8 \\ 0 & -0.8 & 1 \end{bmatrix}$$

$$Q(x_1, x_2, x_3, a) = a \cdot (x_1/4 - x_2 - x_3)$$

$$a \in \{-1, 1\}$$

How bad can it be?

- Pretty bad.
- For, $\mathbf{x}_o = \emptyset$
  we have values
  $\approx (0.18, 0.16, 0.16)$

$$P(X_1, X_2, X_3) \sim \mathcal{N}(\mu, \Sigma)$$

$$\mu = (0,0,0)^\mathsf{T}$$

$$\Sigma = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -0.8 \\ 0 & -0.8 & 1 \end{bmatrix}$$

$$Q(x_1, x_2, x_3, a) = a \cdot (x_1/4 - x_2 - x_3)$$

$$a \in \{-1, 1\}$$

How bad can it be?

- Pretty bad.
- For, $\mathbf{x}_o = \emptyset$
  we have values
  $\approx (0.18, 0.16, 0.16)$
- So reveal *x1*

$$P(X_1, X_2, X_3) \sim \mathcal{N}(\mu, \Sigma)$$

$$\mu = (0, 0, 0)^{\mathsf{T}}$$

$$\Sigma = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -0.8 \\ 0 & -0.8 & 1 \end{bmatrix}$$

$$Q(x_1, x_2, x_3, a) = a \cdot (x_1/4 - x_2 - x_3)$$

$$a \in \{-1, 1\}$$

How bad can it be?

- Pretty bad.
- For, $\mathbf{x}_o = \emptyset$ we have values $\approx (0.18, 0.16, 0.16)$
- So reveal x_1

$$P(X_1, X_2, X_3) \sim \mathcal{N}(\mu, \Sigma)$$

$$\mu = (0,0,0)^\mathsf{T}$$

$$\Sigma = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -0.8 \\ 0 & -0.8 & 1 \end{bmatrix}$$

$$Q(x_1, x_2, x_3, a) = a \cdot (x_1/4 - x_2 - x_3)$$

$$a \in \{-1, 1\}$$

- Then select x_2 or x_3. Value of this policy is about **0.25**.

How bad can it be?

- Pretty bad.
- For, $\mathbf{x}_o = \emptyset$ we have values $\approx (0.18, 0.16, 0.16)$
- So reveal x_1

$$P(X_1, X_2, X_3) \sim \mathcal{N}(\mu, \Sigma)$$

$$\mu = (0, 0, 0)^\mathsf{T}$$

$$\Sigma = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -0.8 \\ 0 & -0.8 & 1 \end{bmatrix}$$

$$Q(x_1, x_2, x_3, a) = a \cdot (x_1/4 - x_2 - x_3)$$

$$a \in \{-1, 1\}$$

- Then select x_2 or x_3. Value of this policy is about **0.25**.
- **Value of "purchase x_2 and x_3" is 0.5**

Where do we go from here?

- Heuristics / Modified Greedy?
  - Most promising: Consider simultaneously revealing *sets* of features. (Solves previous example.)
- Go after the optimal purchasing policy using RL
- In both cases, will want to leverage problem-specific structure

Summary

- Inroduced **Active Action Choice**
  - Related to “Active Learning,” “Budgeted Learning,” “Active Diagnosis”, ...
- Shown that this problem is not submodular, cannot get the (1-1/e) greedy approximation bound
- In light of this, suggested avenues for policies that avoid the greedy catastrophe

References

- K. Deng, J. Pineau and S.A.Murphy (2011). Active Learning for Developing Personalized Treatment. Proceedings of the Twenty-Seventh Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-11). AUAI Press 161-8. Presentation slides.
- K. Deng, J. Pineau and S.A.Murphy (2011). Active Learning for Personalizing Treatment.Adaptive Dynamic Programming And Reinforcement Learning (ADPRL), 2011 IEEE Symposium on. 11-15 April 2011, pages 32-39. Presentation slides.
- Daniel Golovin, Andreas Krause, "Adaptive Submodularity: Theory and Applications in Active Learning and Stochastic Optimization",In Journal of Artificial Intelligence Research (JAIR), vol. 42, pp. 427-486, 2011.
- A. Kapoor, R. Greiner. "Reinforcement Learning for Active Model Selection". Utility-Based Data Mining (UBDM), August 2005.
- Daniel Lizotte, Omid Madani, and Russell Greiner. Budgeted learning of naïve-Bayes classifiers. In 19th Conference on Uncertainty in Artificial Intelligence (UAI), 2003.
- G. L. Nemhauser, L. A. Wolsey and M. L. Fisher. An analysis of approximations for maximizing submodular set functions I, Mathematical Programming 14 (1978), 265–294