15-780 - GRADUATE ARTIFICIAL INTELLIGENCE

AI AND EDUCATION III

Shayan Doroudi

May 1, 2017

Series on applications of AI to education.

Lecture	Application	AI Topics
4/24/17	Learning	Machine Learning + Search
4/26/17	Assessment	Machine Learning + Mechanism Design
5/01/17	Instruction	Multi-Armed Bandits

Prediction

- Predicting performance in a learning environment
- Predicting performance on a test

Intervention

Prediction

- Predicting performance in a learning environment
- Predicting performance on a test

Intervention

- Changing instruction based on refined cognitive model

Prediction

- Predicting performance in a learning environment
- Predicting performance on a test

Intervention

- Changing instruction based on refined cognitive model
- Computerized Adaptive Testing

PREDICTION VS. INTERVENTION

Prediction

- Predicting performance in a learning environment
- Predicting performance on a test

Intervention

- Changing instruction based on refined cognitive model
- Computerized Adaptive Testing

- Choosing the best instruction

- Recall the Randomized Weighted Majority Algorithm.

- Recall the Randomized Weighted Majority Algorithm.
- After each decision, we know if each expert got it right or wrong.

- Recall the Randomized Weighted Majority Algorithm.
- After each decision, we know if each expert got it right or wrong.
- Multi-Armed Bandits: Choose only one arm (expert/action); only know if that arm was good or bad.

- Set of K actions $\mathcal{A} = \{a_1, \ldots, a_K\}$.

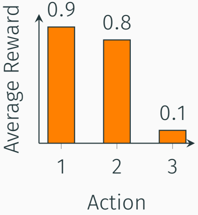
- Set of K actions $\mathcal{A} = \{a_1, \ldots, a_K\}$.
- At each time step t, we choose one action $a_t \in \mathcal{A}$.

- Set of K actions $\mathcal{A} = \{a_1, \ldots, a_K\}$.
- At each time step t, we choose one action $a_t \in \mathcal{A}$.
- Observe reward for that action, coming from some unknown distribution with mean μ_a.

- Set of K actions $\mathcal{A} = \{a_1, \ldots, a_K\}$.
- At each time step t, we choose one action $a_t \in \mathcal{A}$.
- Observe reward for that action, coming from some unknown distribution with mean μ_a.
- Want to minimize regret:

$$R(T) = T \max_{a \in \mathcal{A}} \mu_a - \mathbb{E}\left[\sum_{t=1}^{T} \mu_{a_t}\right]$$

POLL (MULTI-ARMED BANDITS)



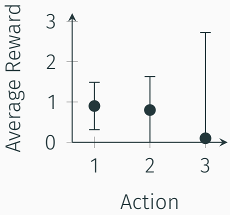
Suppose action 1 was taken 20 times, action 2 was taken 10 times, and action 3 was taken once. Which action should we take next?

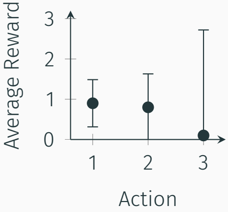
- Action 1
- Action 2
- Action 3
- Some distribution over the actions.

- **Exploration**: Trying different actions to discover what's good.
- **Exploitation**: Doing (exploiting) what we believe to be best.

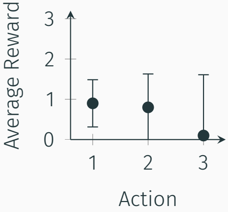
- Explore-then-Commit: Take each action n times, then commit to the action with the best sample average reward.

Upper Confidence Bound (UCB)


Average Reward
3
2
1
0
1
2
3
Action



After taking action 3 two more times and seeing 0.1 both times:



UCB1 Algorithm:

1. Take each action once.
2. Take action

$$\arg\max_{a_j \in \mathcal{A}} \frac{1}{n_j} \sum_{i=1}^{n_j} r_{j,i} + \sqrt{\frac{2\ln(n)}{n_j}}$$

- n is the total number of actions taken so far
- n_j is the number of times we took a_j
- $r_{j,i}$ is the reward from the ith time we took a_j

THOMPSON SAMPLING

Thompson Sampling Algorithm: Choose actions according to the probability that we think they are best.

Thompson Sampling Algorithm: Choose actions according to the probability that we think they are best.

- Take action a_j with probability

$$\int \mathbb{I}(\mathbb{E}\left[r|a_j,\theta\right] = \max_{a\in\mathcal{A}} \mathbb{E}\left[r|a,\theta\right])P(\theta|\mathcal{D})d\theta$$

Thompson Sampling Algorithm: Choose actions according to the probability that we think they are best.

- Take action a_j with probability

$$\int \mathbb{I}(\mathbb{E}\left[r|a_j, \theta\right] = \max_{a \in \mathcal{A}} \mathbb{E}\left[r|a, \theta\right]) P(\theta|\mathcal{D}) d\theta$$

- Can just sample θ according to $P(\theta|\mathcal{D})$, and take $\max_{a \in \mathcal{A}} \mathbb{E}\left[r|a, \theta\right]$

- Suppose each action a_j gives rewards according to a Bernoulli distribution with some unknown probability p_j.

- Suppose each action a_j gives rewards according to a Bernoulli distribution with some unknown probability p_j.
- Use Conjugate Prior (Beta Distribution):

$$P(p_j|\alpha, \beta) \propto p_j^{\alpha}(1 - p_j)^{\beta}$$

- Suppose each action a_j gives rewards according to a Bernoulli distribution with some unknown probability p_j.
- Use Conjugate Prior (Beta Distribution):

$$P(p_j|\alpha, \beta) \propto p_j^{\alpha}(1 - p_j)^{\beta}$$

- After we take a_j, if we see reward r_j,

$$P(p_j|\alpha, \beta, r_j) \propto P(p_j|\alpha, \beta)P(r_j|p_j) \propto p_j^{\alpha}(1 - p_j)^{\beta} p_j^{r_j}(1 - p_j)^{1-r_j}$$

- Suppose each action a_j gives rewards according to a Bernoulli distribution with some unknown probability p_j.
- Use Conjugate Prior (Beta Distribution):

$$P(p_j|\alpha, \beta) \propto p_j^{\alpha}(1-p_j)^{\beta}$$

- After we take a_j, if we see reward r_j,

$$P(p_j|\alpha, \beta, r_j) \propto P(p_j|\alpha, \beta)P(r_j|p_j) \propto p_j^{\alpha}(1-p_j)^{\beta} p_j^{r_j}(1-p_j)^{1-r_j}$$

$$P(p_j|\alpha, \beta, r_j) \propto p_j^{\alpha+r_j}(1-p_j)^{\beta+1-r_j}$$

THOMPSON SAMPLING WITH BETA PRIOR

- Suppose each action a_j gives rewards according to a Bernoulli distribution with some unknown probability p_j.
- Use Conjugate Prior (Beta Distribution):

$$P(p_j|\alpha, \beta) \propto p_j^{\alpha}(1 - p_j)^{\beta}$$

- After we take a_j, if we see reward r_j,

$$P(p_j|\alpha, \beta, r_j) \propto P(p_j|\alpha, \beta)P(r_j|p_j) \propto p_j^{\alpha}(1 - p_j)^{\beta} p_j^{r_j}(1 - p_j)^{1-r_j}$$

$$P(p_j|\alpha, \beta, r_j) \propto p_j^{\alpha+r_j}(1 - p_j)^{\beta+1-r_j}$$

- After any action the posterior distribution will be as follows:

$$P(p_j|\mathcal{D}) \propto p_j^{\alpha+s_j}(1 - p_j)^{\beta+f_j}$$

Thompson Sampling Algorithm with Bernoulli Actions and Beta Prior:

- Sample $p_1, \ldots, p_K$ with probability

$$P(p_j|\mathcal{D}) \propto p_j^{\alpha+s_j}(1-p_j)^{\beta+f_j}$$

- Choose $\arg\max_{a_j \in \mathcal{A}} \mathbb{E}\left[r|p_j\right] = p_j$

How can we increase exploration using Thompson Sampling with Beta Prior?

- Choose a large α
- Choose a large β
- Choose an equally large α and β
- Beats me

AXIS: Generating Explanations at Scale with Learnersourcing and Machine Learning

Joseph Jay Williams[1] **Juho Kim**[2] **Anna Rafferty**[3] **Samuel Maldonado**[4]
Krzysztof Z. Gajos[1] **Walter S. Lasecki**[5] **Neil Heffernan**[4]

[1]Harvard University
Cambridge, MA
joseph_jay_williams@harvard.edu, kgajos@eecs.harvard.edu

[2]Stanford University & KAIST
Stanford, CA
juhokim@cs.kaist.ac.kr

[3]Carleton College
Northfield, MN
arafferty@carleton.edu

[4]WPI
Worcester, PA
{sjmaldonado,nth}@wpi.edu

[5]Computer Science & Engineering
University of Michigan, Ann Arbor
wlasecki@umich.edu

EXAMPLE: AXIS

Chris has a cookie jar that contains 5 chocolate cookies, and 3 oatmeal cookies. He will draw two cookies from the jar one at a time without replacing the first cookie.

What is the probability that Chris gets a chocolate cookie on his first draw and an oatmeal cookie on his second draw?

Enter your answer below:

Explanation: Here is an explanation someone wrote of why the answer is right, and how to solve the problem.

The probability of getting a chocolate cookie on his first draw is 5/8. If he draws a chocolate cookie, there will be 4 chocolate cookies and 3 oatmeal cookies left, so the probability of getting an oatmeal cookie on his second draw is 3/7. (5/8)*(3/7)=15/56.

How helpful do you think this explanation is for learning?

Absolutely Unhelpful									Perfect
1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○

EXAMPLE: AXIS

	Discarded Learner Explanations (removed by AXIS Filtering Rule)	AXIS 75: Presented by AXIS after interacting with 75 learners	AXIS 150: Presented by AXIS after interacting with 75 learners	Instructional Designer's Explanations	Practice Problems Only (No Explanations)
Explanation Rating (1-Unhelpful to 10-Excellent)	6.03 (3.01)	6.57 (2.84)	6.83 (2.45)	7.30 (2.45)	---
Increase in Perceived Likelihood of Solving Problem (1 – 10 Scale)	0.69 (2.78)	0.57 (2.66)	0.71 (2.71)	0.48 (2.51)	-0.01 (2.30)
Accuracy Increase in Solving Problems	0.02 (0.47)	0.12 (0.44)	0.12 (0.46)	0.09 (0.47)	0.03 (0.48)
Accuracy Increase: Problems Isomorphic to Study	0.19 (0.60)	0.23 (0.52)	0.23 (0.55)	0.17 (0.57)	0.16 (0.58)
Accuracy Increase: Transfer Problems	-0.06 (0.49)	0.06 (0.48)	0.07 (0.50)	0.05 (0.51)	-0.04 (0.51)

What's missing?

CONTEXTUAL BANDITS

- Obtain some context $x_{t,a}$
- Assume linear payoff function:

$$\mathbb{E}\left[r_{t,a}|x_{t,a}\right] = x_t^T\theta_a$$

- Obtain some context $x_{t,a}$
- Assume linear payoff function:

$$\mathbb{E}\left[r_{t,a}|x_{t,a}\right] = x_t^T\theta_a$$

- Obtain some context $x_{t,a}$
- Assume linear payoff function:

$$\mathbb{E}\left[r_{t,a}|x_{t,a}\right] = x_t^T \theta_a$$

- Solve for θ_a using linear regression, build confidence intervals over the mean, and apply UCB.

Thompson Sampling Algorithm with Context:

- Get context x
- Take action a_j with probability

$$\int \mathbb{I}(\mathbb{E}\left[r|x, a_j, \theta\right] = \max_{a \in \mathcal{A}} \mathbb{E}\left[r|x, a, \theta\right]) P(\theta|\mathcal{D}) d\theta$$

- Can just sample θ according to $P(\theta|\mathcal{D})$, and take $\max_{a \in \mathcal{A}} \mathbb{E}\left[r|x, a, \theta\right]$

- Multi-armed bandits can help decide what instructional activities to give to students.
- Saw a frequentist (UCB) and Bayesian (Thompson Sampling) algorithm for multi-armed bandits.
- Contextual bandits can help personalize decisions for students and reinforcement learning can help make adaptive decisions for students.