Connecting Designers, Behavioral Scientists, and Reinforcement Learning Researchers via Collaborative, Dynamic, Personalized A/B Experimentation

Joseph Jay Williams williams@comp.nus.edu.sg www.josephjaywilliams.com
School of Computing, Department of Information Systems & Analytics, NUS-HCI Lab

We reimagine randomized A/B experiments in digital educational resources as a collaborative tool, helping instructors discover which lessons help students, enabling learning scientists to investigate psychological theories, and providing a real-world test bed for algorithms developed by reinforcement learning researchers. We provide the MOOClet/AdapComp software requirements specification for online experiments, which provides an abstraction for any reinforcement learning algorithm to adapt experiments. Machine learning researchers can use our system's API to evaluate their algorithms' policies for adaptive experiments, dynamically trading off exploration and exploitation to present the best experimental conditions to future learners, and personalizing by delivering alternative conditions based on characteristics of a learner.

Multi-Armed Bandits for Crowdsourced, Dynamic, Experimentation: Explanations for Math Problems

HOW DO WE ENHANCE STUDENTS' LEARNING FROM ONLINE MATH PROBLEMS?

Mary's music store had 5 truck loads of CDs delivered. Each truck dropped off 12 boxes. Each box has ***c*** CDs. Write an expression for how many CDs were delivered.

The **correct answer** is 60c

Explanation for answer:
Take the problem step by step.
Every truck has 12 boxes and there are 5 trucks, so how many boxes are there?
12*5 = 60.

How helpful was this explanation?
completely unhelpful — extremely helpful
0 1 2 3 4 5 6 7 8 9 10

Action a $\in A$

REWARD *R*

Randomized Probability Matching, Thompson Sampling

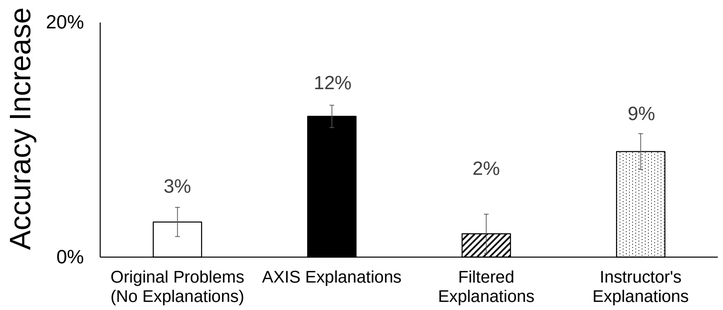
(e.g. Chapelle & Li, 2013)

Policy π Parameters θ

$\epsilon_i \sim Beta(\alpha, \beta)$ (Probability of Explanation being Rated Helpful)

$R \sim Bin(10, \epsilon_i)$ (0 to 10 Rating by Student)

$P(\theta|D) \propto \prod P(r_i|a_i, x_i, \theta)P(\theta)$



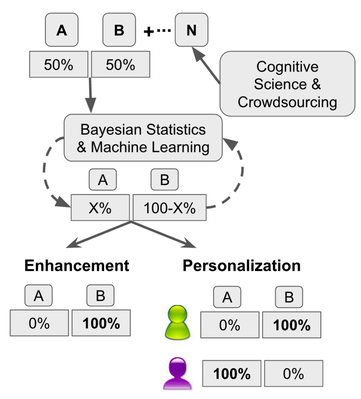
Test bed for RL algorithms: Policies for Adaptive Experimentation

When A/B experiments are implemented using the MOOClet/AdapComp Software Requirements Specification:

- Researchers can use the API to access data about past rewards from taking actions, & learners' context.
- Researchers can dynamically modify the policy for assigning learners to experimental conditions.

API Endpoints (MOOClet/AdapComp Specification)

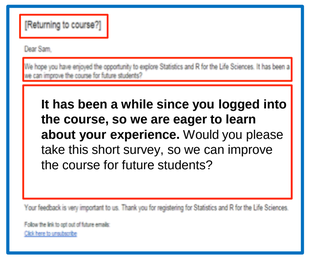
Endpoint	Parameters	Returns
getLearnerContext	learner_id	{age: 28, days_active: 2, ...}
getPastRewards	adapcomp_id	{learner_id1: reward_value, learnerid2: ...}
assignLearnerCondition	learner_id, adapcomp_id, condition	{learner_id, adapcomp_id, condition}



Dynamic, Personalized Experimentation: Motivational Emails

HOW TO MOTIVATE PEOPLE OF DIFFERENT AGES & ACTIVITY LEVELS TO RESPOND TO EMAILS?

Actions:
3 Subject Lines
x 3 Intro Messages
x 3 Email Body
= 27 Versions


[Returning to course?]

Dear Sam,

We hope you have enjoyed the opportunity to explore Statistics and R for the Life Sciences. It has been a while since you logged into the course, so we can improve the course for future students?

It has been a while since you logged into the course, so we are eager to learn about your experience. Would you please take this short survey, so we can improve the course for future students?

Your feedback is very important to us. Thank you for registering for Statistics and R for the Life Sciences.

Follow the link to opt out of future emails
Click here to unsubscribe

REWARD *R*

Responded to email [0, 1]

CONTEXT:
Age Group = [18-22, 23-26, 27-35, >36]

Number of Days Active = [0, 1, >2]

First batch of 4000 people– Proportion getting Version A, B or C

Age	A	B	C
18-22	0.33	0.33	0.33
23-26	0.33	0.33	0.33

3.8%
3.8% + 1.6% + 2.3%
=0.48

SECOND batch of 1500 people– Proportion getting Version A, B or C

Age	A	B	C
18-22	0.48	0.22	0.30
23-26	0.60	0.12	0.28

Percentage responding to email

	A	B	C
18-22	3.8%	1.6%	2.3%
23-26	5.0%	1.0%	2.3%

Random Assignment	Weighted Personalization	Difference	Percentage Increase
4.5%	7.2%	2.7%	60%