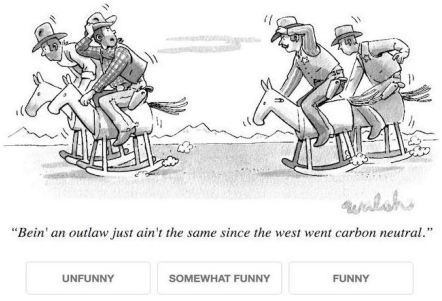
I am a machine learning researcher working on sequential decision-making in feedback loops, i.e., the multi-armed bandit problem. With backgrounds on online optimization and interdisciplinary collaborations, I develop algorithms with theoretical guarantees that have important real-world applications.

Modern practitioners use machine learning to guide their decisions in applications ranging from marketing and targeting to policy making and clinical trials. To embrace automated decision-making in closed loops, it is crucial to make decisions safely and efficiently, as poor decisions waste human resources and result in social and economic costs. To this end, I propose to build a comprehensive research program on closed-loop decision problems with an emphasis on humans in the loop. Specifically, I will investigate novel ways of interacting with humans in realistic settings, develop mathematical algorithms and theory for complex models, and apply them to salient problems in biology, psychology, and economics.

Learning to make better decisions in feedback loops is challenging. Consider a personalized recommendation system. The system decides which item to show to a user and then, based on the click feedback, updates its model. However, the clicks are only available for items chosen by the previous model. This *feedback loop* introduces bias in the collected data. As a result, a system that naïvely chooses the best item from the updated model is known to get stuck in recommending suboptimal items. The cost of such a failure would be significant if the task was clinical trials for recommending various treatments. Despite recent developments such as deployable web decision services [1,2], near-optimal sequential decision-making beyond simple problems remains open, especially with complex models such as deep neural networks.

The New Yorker Cartoon Caption Contest, as another example, uses a crowdsourcing system to evaluate thousands of submitted captions and pick the funniest. Upon serving a crowdworker, the system has to decide which caption needs to be evaluated by the worker. Given limited workers, we hope to quickly rule out unfunny captions as we want more ratings (thus more confidence) on the top contenders. How can we do this without accidentally ruling out the funniest one? Can we instead use comparative judgments? How about cartoon/caption-related features?



A caption being evaluated by a worker.

Learning to make decisions in complex environments with humans gives rise to diverse research opportunities, requiring not only skills in probability, bandit theory, and online optimization, but also experience in interdisciplinary collaborations. As these are my areas of expertise, I believe I am uniquely well-suited for this pursuit. My expertise is exemplified by:

- Personalized recommendations at scale by combining bandits with online algorithms and hashings [NIPS17].
- Adaptive sampling algorithms with applications in sample-efficient planning for biology experiments and cartoon caption contests [ICML16,AISTATS16].
- Collaborations with psychologists, developing probabilistic models for understanding human memory search and classifying brain-damaged patients [NIPS15,ICML13].
- Online optimization for tracking and adapting to nonstationary environments [AISTATS17].

I will now describe each project in detail, followed by a discussion of planned future work.

Interactive Search at Scale

Consider the following interactive search scenario. The system (the "learner") retrieves an item ("pulls an arm") for a user and receives a relevance score (a "reward") from her, repeating the process until she is satisfied. The system is evaluated by the cumulative score. This is an instance of the Multi-Armed Bandit (MAB) problem; MAB jargons are used in parentheses above. MAB is a state-less version of reinforcement learning but enjoys stronger theoretical guarantees in general. Since the feedback is given just for the items shown to the user, an algorithm faces the explore-exploit dilemma. That is, retrieving various items ("exploration") reveals new information but not rewards, and a strategy of exploiting imperfect information for immediate rewards could get stuck with suboptimal arms.

A popular variant of MAB is Generalized Linear Bandit (GLB) that uses arm features to maximize rewards with significantly fewer samples. Existing GLBs are, however, not scalable with the number of iterations and the number of items in the database, which blocks its practical usage. In [NIPS17], we resolved these issues via online parameter updates and a novel combination with hashings, allowing GLBs to be deployed at practical scale. The key is a novel reduction framework that takes *any* online learning algorithm and turns it into a bandit algorithm, achieving both scalability and generality. Equipped with state-of-the-art theoretical bounds, our scalable methods empirically runs much faster than prior work without sacrificing the statistical efficiency.

Adaptive Biology Experiments

Consider the following biology experiment. A virologist wants to find the top-k genes (arms) that maximally affect virus replication. A unit experiment (an arm pull) is testing a gene with a virus and observing the amount of replication (reward). As the observations are noisy, one should perform repeated experiments to obtain stable numbers. How can we save the experiment budget while maximizing the accuracy of identifying the right top-k genes? A naïve solution of measuring each gene an equal number of times wastes a lot of experiment budget on genes that are not interesting. This is another variant of MAB that I made a number of contributions.

For a realistic setting where one can perform multiple experiments at a time (e.g., a microwell array can test 400 genes at once), we proposed novel algorithms that fully utilize the structure [AISTATS16]. In real-world biology experiments and social media monitoring tasks, my methods identify the top-k arms much faster than prior work that ignores the structure. Furthermore, I noticed that in practice the sampling budget is often limited and unknown; e.g., in cartoon caption contests, the number of participating crowdworkers is not known a priori. Departing from prior work, our proposed method is adaptive to the unknown budget, enjoying a theoretical guarantee for *any* budget [ICML16]. Ours show better sample efficiency than nonadaptive ones in cartoon caption contest tasks.

Understanding Human Memory Search for Psychology

Psychologists are interested in understanding how humans search through their memory. One popular task they consider is to ask a human to name items belonging to a given category; e.g., "Name as many animals as possible in 60 seconds." The output list of items is known to contain rich information about human brains. Developing machine learning models for such a list generation process can not only help understand human brains but also be used to classify brain-damaged patients from healthy people. However, the modeling is challenging as the lists do not fit the traditional paradigm of sampling with or without replacement.

To this end, I performed fruitful collaborations with psychologists and developed new models and inference methods for capturing rich information like item importance and item similarities. Our models are successfully combined with downstream procedures for classifying brain-damaged patients and building document classifiers without labeled data [ICML13]. Furthermore, our new random walk model efficiently estimates the underlying network of items, providing a great tool for understanding and visualizing human memory search for psychologists [NIPS15,CogSci16].

Online Learning under Changing Environments

Concept drift is abundant. Consider that we have a trained model for classifying which presidential candidate a social media post is supporting. As the words describing presidential candidates change as new issues come up, the performance of the classifier decays over time without proper retraining by engineers. How can we build a learner that automatically adapts to changing environments without human interventions?

In [AISTATS17], we developed an efficient framework that turns *any* online learning algorithm into the one that adapts to changing environments. Such a reduction framework is powerful as it *(i)* works for all online learning tasks rather than for a particular task and *(ii)* results in a strongly adaptive algorithm that does not need to know *when* or *how frequently* the environment changes. Equipped with improved theoretical bounds from prior work, our algorithm is able to catch up the environmental change much faster than existing work in an empirical study. The bound is even more improved in a journal extension of the same work [J17] (under review).

Future Directions

Sample-Efficient Decision-Makings with Humans Despite the success of MABs in applications like personalized recommendations, one serious drawback is the slow convergence to the optimal decision. The issue is usually not the CPU cycles but rather the number of human feedbacks that are more costly. To this end, I aim to design new types of feedback that require minimal human efforts but result in faster convergence. I propose two directions. **First**, imagine we ask users not only for relevance feedback but also *feature* feedback. For document recommendations, for example, we may ask users to optionally click on words that are particularly relevant. For images, we can ask users to draw a bounding box indicating the most relevant region. **Second**, imagine that the users are prompted occasionally to answer a comparative judgment question such as "is A closer to B or C?". A recent study shows that such feedback efficiently helps find a low-dimensional representation [3]. What is the best way to combine the relevance feedback with comparative judgments? I believe the background on MAB and human response modeling puts me in an excellent position to answer the question.

Understanding Human Decision-Making The explore-exploit dilemma in MAB problems is also encountered by humans in daily life. For example, humans explore different locations to live and decide to settle down at some point. Can we better understand how humans make decisions through the lens of MABs? Given a data set of sequential human decisions, can we reverse-engineer the human decision-making process? A solution to this question is intriguing to psychologists and economists, providing a tool for understanding human decisions, performing counterfactual analyses, and guiding interventions to help humans. Despite some preliminary attempts, there are no principled ways of estimating the learner's decision-making process in MABs. I am currently working on this problem with a colleague in the economics department. My experience in interdisciplinary work gives me the right background to attack this problem.

References

[NIPS17]
Kwang-Sung Jun, Aniruddha Bhargava, Robert Nowak, and Rebecca Willett.
"Scalable Generalized Linear Bandits: Online Computation and Hashing".
In *Neural Information Processing Systems* (NIPS), 2017.

[AISTATS17]
Kwang-Sung Jun, Francesco Orabona, Rebecca Willett, and Stephen Wright.
"Improved Strongly Adaptive Online Learning using Coin Betting".
In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2017. **Oral presentation**.

[J17]
Kwang-Sung Jun, Francesco Orabona, Rebecca Willett, and Stephen Wright.
"Online Learning for Changing Environments using Coin Betting".
arXiv:1711.02545, 2017.

[CogSci16]
Jeffrey Zemla, Yoed Kenett, **Kwang-Sung Jun**, and Joseph Austerweil.
"U-INVITE: Estimating Individual Semantic Networks from Fluency Data".
In *Proceedings of the Annual Meeting of the Cognitive Science Society (CogSci)*, 2016.

[ICML16]
Kwang-Sung Jun and Robert Nowak.
"Anytime Exploration for Multi-armed Bandits using Confidence Information".
In *International Conference on Machine Learning (ICML)*, 2016.

[AISTATS16]
Kwang-Sung Jun, Kevin Jamieson, Rob Nowak, and Xiaojin Zhu.
"Top arm identification in multi-armed bandits with batch arm pulls."
In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2016.

[NIPS15]
Kwang-Sung Jun, Xiaojin Zhu, Timothy Rogers, Zhuoran Yang, and Ming Yuan.
"Human memory search as initial-visit emitting random walk".
In *Neural Information Processing Systems (NIPS)*, 2015.

[ICML13]
Kwang-Sung Jun, Xiaojin Zhu, Burr Settles, and Timothy Rogers.
"Learning from Human-Generated Lists."
In *International Conference on Machine Learning (ICML)*, 2013.

[1] Alekh Agarwal, Sarah Bird, Markus Cozowicz, Luong Hoang, John Langford, Stephen Lee, Jiaji Li, Dan Melamed, Gal Oshri, Oswaldo Ribas, Siddhartha Sen, Alex Slivkins. "A Multiworld Testing Decision Service". arXiv:1606.03966, 2016.

[2] `https://mwtds.azurewebsites.net/`

[3] Lalit Jain, Blake Mason, Robert Nowak. "Learning Low-Dimensional Metrics". In arXiv:1709.06171, 2017.