Machine Learning (ML) has played a transformational role in engineering and science. Nowadays, practitioners make decisions by ML in applications ranging from product recommendation to scientific discovery. However, most traditional models are built on top of the idealized *unbiased* data setting (i.e., i.i.d.), which is often far from reality. This mismatch can lead to suboptimal decisions. For example, a personalized recommendation model is trained with its own user clicks, but the clicks are available only for those items chosen by the previous model. This feedback-loop introduces bias, and models assuming unbiased data make suboptimal decisions. How can we handle this bias while converging to the optimal decision? Note that the feedback is often from humans. Can we optimally interact with humans to accelerate the convergence? Another prevalent non-i.i.d. case is where trends or concepts change over time (e.g., the stock market, political issues, etc.). How can a model quickly adapt to changing environments? These challenges are crucial for building efficient closed-loop, adaptive decision systems that can solve multidisciplinary problems, giving rise to diverse research opportunities.

Research Summary I focus on (A) ***learning with limited feedback (multi-armed bandit)*** and also work on (B) cognitive modeling and (C) learning under nonstationary environments. For these problems, I developed novel methods that excel in both theoretical guarantees and real-world tasks, relying on tools from statistics, optimization, and learning theory. My close collaboration with psychologists led me to understand that the tight connection between theory and practice is the key to impactful research.

(A) Learning with Limited Feedback Multi-Armed Bandit (**MAB**) is a state-less version of Reinforcement Learning. In MAB, a learner makes repeated decisions in a feedback-loop. This feedback gives rise to the explore-exploit dilemma. Here, my contribution is two-fold. *First*, in a generalized linear variant, I resolved scalability issues that hinder practical implementations in, for example, personalized recommendation systems. My contribution involves new efficient online algorithms and a novel combination with hashings. *Second*, I proposed new algorithms for identifying top-k decisions in *realistic* settings: (i) making multiple decisions at once (ii) dealing with unknown sampling budgets. These methods were able to reduce the costs of adaptive biology experiments and cartoon caption contests.

(B) Human Memory Search Consider the following task: "Name as many animals as possible in 60 seconds." The response is a list that carries crucial information about animals and your brain, but its non-i.i.d. nature poses modeling challenges. I developed new models and inference methods that capture rich information like item importance and item similarities. These are successfully combined with downstream procedures for *(i)* classifying brain-damaged patients, *(ii)* analyzing the structure of human memory search, and *(iii)* building document classifiers without labeled data. This was a fruitful collaboration with psychologists, resulting in a versatile model for both understanding humans and building classifiers.

(C) Tracking Changing Environments Imagine you wish to track popular topics in social media that change over time. How can we adapt to the changing environment without knowing explicitly *when* and *how frequently* the changes happen? I developed an efficient framework that turns *any* online learning algorithm into one that adapts to changes. I was able to theoretically analyze this method and showed that it adapts to *any* number of environmental changes with no prior knowledge. The proposed algorithm improves upon the state-of-the-art in terms of both the theoretical bound and the empirical evaluation.

Future Directions Revolving around decision making under limited-feedback environments, my research program opens up exciting opportunities in theory, algorithms, and applications in other disciplines. My long-term goal is to build a healthy research ecosystem by maintaining an open source decision framework, which will accelerate collaborations and identify new problems. My two short-term directions are as follows: *(i)* How can we speed up the convergence of MABs? The iterations of MABs costs money and/or human resources rather than CPU cycles. Specific directions include not only exploiting the model structure, but also studying what kind of low-cost-but-informative additional feedback we can get from humans. *(ii)* How can we study the human decision making through the lens of MABs? I will focus on efficient and principled ways of inferring the human decision process with MABs, which will be useful in other disciplines like economics and psychology.

In the remainder of the statement, I elaborate the above topics.

1 Multi-Armed Bandits (MABs)

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 relevance score. This is an instance of the Multi-Armed Bandit (MAB) problem; MAB jargons are used in parentheses above. Since the feedback is given for the items shown to the user only, 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 to myopically get more reward could repeatedly pull suboptimal arms.

When given item-specific features, one can use them to posit a generalized linear model for the reward generation. Existing generalized linear bandits are, however, not scalable with the number of iterations and the number of items in the database, which blocks its practical usage. In [NIPS17], I developed a scalable solution through online parameter updates, which is an order of magnitude faster than prior art. The core technique is a novel reduction framework that takes *any* online learning algorithm and turns it into a bandit algorithm, achieving both scalability and generality. I further achieve a sublinear time complexity in the number of items via a novel combination of locality-sensitive hashings. Equipped with state-of-the-art theoretical bounds, our scalable methods empirically show competitive statistical efficiency to existing non-scalable methods. As a result of this work, generalized linear bandits can be deployed at practical scale.

Finding the Top-k Arms As a variant of MAB, consider the following biological experiments. 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) with some noise. This "covariate-less" experimental design is challenging since naïvely measuring each gene an equal number of times wastes a lot of measurement budget. Efficient methods must adaptively focus on experimenting the genes that are near the k-th best gene. In [AISTATS16], I consider the setting where one can perform b experiments at a time; e.g., a microwell array can test 400 genes at once. I provide theoretical analysis on how many microwell arrays are needed to accurately identify the top-k genes. The evaluations on real-world biology experiments and social media monitoring tasks show that our methods identify the top-k arms much faster than prior work. In [ICML16], I notice that in real world the sampling budget is often limited and unknown. For example, in the New Yorker cartoon caption contest, captions are evaluated via crowdsourcing where the number of participants 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. Ours show better sample efficiency than nonadaptive ones in cartoon caption contest tasks.

2 Human Memory Search for both Machine Learning and Psychology

Human-generated lists, a list of items belonging to a category, are a form of non-i.i.d. data with applications in machine learning and psychology. The modeling challenge of such a list generation is that the lists do not fit the traditional sampling paradigm with or without replacement. In [ICML13], I developed a new model called Sampling WIth Reduced repLacement (SWIRL) that sits in between the sampling with and without replacement. Equipped with fast inference methods, SWIRL successfully captures rich information like item importance, the degree of repeat, and the length distribution, all at once. In an ML application, I use the item importance to successfully construct classifiers with no labeled documents ("zero-shot"), enjoying better accuracies than competitors. In a clinical application, the learned model is successfully turned into a classifier telling whether a given list is from a brain-damaged patient or not.

In [NIPS15], I developed a random walk model that captures the item similarities. While a similar model was considered by psychologists and proved useful for studying human memory search process, there were no tractable parameter estimation methods. I developed an efficient estimation method and proved its statistical consistency. In experiments, our method outperforms baseline methods that are frequently used by psychologists. In a collaboration with psychologists [CogSci16], we further incorporated timing information generated by the participants, which provides invaluable information about item clusters.

All in all, this series of studies became a fruitful collaboration with psychologists and statisticians, resulting in solutions that are interesting for both machine learning and psychology.

3 Online Learning under Changing Environments

Concept drift is abundant. Performance of classifiers for product reviews, fashion items, and political issues decay over time without proper intervention by engineers. How can we build a learner that automatically adapts to changing environments without human interventions? In [AISTATS17], I developed an efficient algorithm 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. At its heart is a novel extension of a parameter-free online learning algorithm. My theoretical analysis shows that the regret bound of the proposed algorithm improves upon the state-of-the-art. The new algorithm is tested in multiple scenarios where we observe that it catches up the environmental change much faster than existing work. My work provides a strong improvement in a different regret bound targeted for 'smoothly-changing' environment, as verified in a recent work [1].

In a journal extension (under review), I propose an improved algorithm that has so-called *data-dependent* regret bound that is adaptive to the "easiness" of the data. This is the first data-dependent regret bound for online convex optimization in changing environments. Experiments show that the new algorithm incurs lower losses than prior work in general.

4 Future Directions

Speeding up the Convergence of MABs Despite the success of MABs in applications like personalized recommendations, one serious drawback is the slow convergence. The issue is often not the computational time but rather the number of human feedbacks or lab experiments that are more costly. I, therefore, aim to design new types of feedback that require minimal human efforts but result in faster convergence. I propose two directions. **First**, consider that we ask users not only for 'relevance' information but also for feedback on 'feature' information. When each instance is a document, for example, we can 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 in the passive setting [2]. In an active setting like MAB, how can we reduce dimensionality with little supervision? 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.

MABs and Human Decision Making The explore-exploit dilemma in MAB problems is also encountered by humans in daily life. Can we better understand how humans make decisions through the lens of MABs? Given a data set of sequential human decisions, how 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 and predicting human decisions, performing counterfactual analysis, and guiding interventions to help humans. For example, how do humans explore different locations and decide to settle down at a location? On what conditions humans learn to make better decisions? 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 with psychologists gives me the right background to attack this problem.

Final Words My current and future research program consists of learning with limited feedback and their interaction with environments and humans. These topics are key components for successful closed-loop systems that learn to make better decisions over time, which provides useful tools for multiple disciplines including biology, psychology, and economics. Through fruitful collaborations with my current and future colleagues with diverse backgrounds, I plan to build a healthy ecosystem of theory, practice, and applications in multidisciplinary problems. Towards this end, I plan to maintain an open source decision framework that nonexperts can easily use and expedites the connection to real-world tasks. Being comprehensive, my program is suited for undergraduates, graduate students of all levels, and postdocs.

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**.

[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] Lijun Zhang, Tianbao Yang, Rong Jin, Zhi-Hua Zhou. "Strongly Adaptive Regret Implies Optimally Dynamic Regret". In arXiv:1701.07570, 2017.

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