We are awash in a deluge of data today. A massive amount of information is being generated at an unprecedented scale in social, financial and scientific domains, and has been utilized to drive nearly every aspect of our modern society. While the promise of big data is significant in some domains, there remain many technical challenges that must be addressed to fully leverage its potential. A crucial bottleneck for data analysis is the *limited capability* of accessing and processing the existing data, due to bandwidth, power, budget and computational constraints. Besides, much of this data might be abundant, and only a small subset can be sufficient for making decisions. This leads to the following fundamental question in artificial intelligence:

How can we intelligently acquire information for decision making, when facing a large volume of data?

Such problems, a.k.a. *optimal information acquisition*, has been extensively studied in many areas, including optimal experimental design, decision theory, operations research, multi-agent systems, sensor networks, and so on. Most existing techniques for addressing such problems assume simple interactions between the learning system and the environment, e.g., by asking for labels (mostly in the form of "yes" or "no") of a query. However, in many real-world applications such as robotics or human-centered decision systems, the learning system and the environment (nature, another learning system, or a human annotator) could communicate through a much richer interface, e.g., in a sequence of actions/feedback, or through the language of rules, advice or explanations, other than simple labels for data points. This raises another fundamental question in AI, especially when building an interactive system:

How can we design a system that learns from complex, structured data, while making interpretable predictions?

Motivated by the above questions, my passion and research interest lie broadly in the design, analysis, and implementation of novel algorithms for probabilistic reasoning, (interactive) machine learning and decision making. More specifically, I am interested in developing resource-efficient, robust, and interpretable learning systems that *actively* extract information, identify the most relevant data for particular learning tasks and make effective decisions under uncertainty. I am also keen on evaluating new algorithmic tools on *real-world applications*, in interdisciplinary domains such as computer vision, program analysis and verification, social network analysis, computational sustainability, machine teaching and human learning, cyber-physical systems, and the physical sciences (e.g., chemical/protein engineering, material science, mechanical engineering).

Prior Work: Optimizing Decision Making

As illustrated in Figure 1, my research has been centered around *optimal decision making* — a paradigm to actively acquire and process information, while minimizing unnecessary interactions with the environment (e.g., a domain expert, a costly information source with limited bandwidth, or a human learner). It captures a broad range of sequential decision making problems arising in artificial intelligence and related fields, including adaptive control, active perception, automated diagnostics, optimal experimental design, intelligent tutoring and adaptive curriculum design, etc. Concretely, my previous research is tightly connected to the following core areas in machine learning:

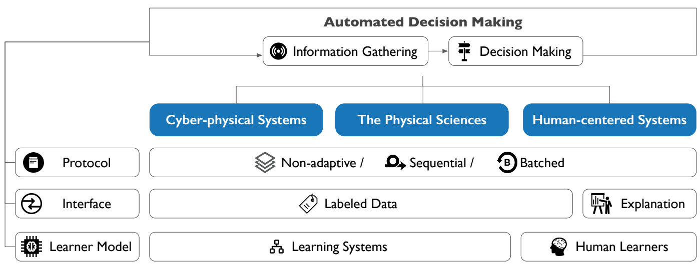

Figure 1: Project Roadmap: Optimizing Decision Making with AI

- *Active learning.* Active learning addresses the problem of selecting an optimal sequence of tests, to determine a hypothesis consistent with the test outcomes. The theoretical aim of active learning is to understand the label complexity, i.e., how many tests are sufficient (and necessary) to learn a hypothesis.
- *Bayesian global optimization.* A fundamental problem in sequential decision making is the exploration-exploitation dilemma: we have to sequentially make decisions with uncertain outcomes, with a potentially large (even infinite) number of available decisions; at each step, we have to trade estimating the function (exploration) and choosing decisions maximizing the estimated function (exploitation).
- *Interpretable models.* Interpretable predictive modeling is useful for very wide application domains where black-box models are not preferred (e.g., medical/biological research, emergency response planning, decision making in financial management, etc). In these scenarios, decisions are critical and often have serious consequences, and domain experts hence would like to know how these predictions are made.
- *Machine teaching.* A natural approach towards interpretable models is to develop algorithms that can properly explain a concept. Unlike active learning, machine teaching involves producing a sequence of examples consistent with a *known* target concept such that the target can be learned by a learner.

Main Research Results

Theoretical Foundations and Practical Challenges of Active Exploration

Theoretical Foundations. In my Ph.D. thesis, I looked into various theoretical variants of the *adaptive information acquisition* problem, where it is challenging to optimize the value of information due to complex constraints and modeling assumptions, such as uncertain inputs, indirect information, delayed feedback in parallel systems, and incomplete knowledge about the model. In a basic (and perhaps the most popular) variant of such problem, the goal is to learn the value of some target random variable through a sequence of conditionally independent, possibly noisy tests. Here, the value of information is defined in terms of the informativeness of the tests performed, measured by Shannon's mutual information. Such information measure has been used in practice since Lindley (1956) [18]; however, little was known about its theoretical behavior—to bridge this gap, we provided the first rigorous analysis of the greedy algorithm under the persistent noise setting [5]. Furthermore, I proved that for a large class of *optimal value of information* problems, one could devise efficient surrogate objectives [7] that are (i) aligned with the original optimization problem, (ii) efficient to evaluate and (iii) approximately *adaptive submodular* [14]. This theoretical insight allowed me and my collaborators to develop principled, efficient and robust adaptive optimization algorithms for a wide variety of sequential problems with strong theoretical guarantees [16, 6, 9]. For example, in a real-world robotic manipulation task, our algorithm achieved competitive empirical performance while being significantly (exponentially in theory) faster than the state-of-the-art baseline [7].

Practical Challenges. Beyond the fundamentals of the optimal value of information problem, I also investigated practical challenges for adaptive information acquisition. For instance, in many application scenarios, fully sequential selection of tests could be infeasible. This motivated me to study *information-parallel active learning* and decision making. I identified a class of batch-mode adaptive stochastic optimization problems, where a simple approach which greedily picked examples in batches and assembled them in a greedy manner, could be competitive with the optimal sequential algorithm [8]. Furthermore, in practice, the underlying probabilistic model defining the value of information may be unknown to the algorithm. This motivated me to study the *online optimal value of information* problem. I proposed a principled algorithmic framework which adapted our offline algorithms to learn the model parameters on the fly, and established a rigorous bound on the expected regret. In collaboration with researchers from Xerox Research Center Europe, we demonstrated the asymmetrically optimal behavior our algorithm on a real-world online interactive troubleshooting task [9].

AI for Experimental Design: Sequential Decision Making and the Exploration-Exploitation Dilemma

Sequential Decision Making with Complex Constraints. At Caltech, I grasped the exciting opportunity to explore interdisciplinary collaborations with domain experts in the physical sciences. Recently, I have established collaborations with biologists and chemists at Caltech, where we devised novel Bayesian active learning and experimental design problems, and developed algorithmic tools for protein/virus engineering. In particular, we identified a novel problem setting that we referred to as *Batched Stochastic Bayesian Optimization*, which captures the experimental setup for a large class of biochemical engineering applications. Our results have demonstrated promising potentials of data-driven approaches in designing high-throughput sequencing experiments [21]. Furthermore, in collaboration with mechanical engineers from the University of Texas, Austin, we have developed computational models of the temperature dynamics of the additive manufacturing system (3D Selective Laser Sintering). This research has set the basis for an ideal test bed for our ongoing effort in the design and evaluation of sequential decision making algorithms for additive manufacturing.

Sequential Decision Making with (Multiple) Computational Resources. Another important aspect for exploration-exploitation is to properly trade off accessing multiple computational resources with different costs. Towards this goal, I collaborate with material scientists at Caltech, where we developed *multi-fidelity Bayesian optimization* strategies for nanophotonics structure discovery [20]. As another example of AI for sequential decision making, I proposed a principled approach to *active object detection* and showed that for a rich class of base detection algorithms, one could derive a natural sequential decision making problem for deciding when to invoke expert supervision (a human export, or an expensive computational oracle). In collaboration with ecologists at ETH, we demonstrated the effectiveness of our algorithm on a biodiversity monitoring application, where it drastically outperformed the state-of-the-art approach, when detecting orangutan nests from a collection of aerial photos taken by conservation drones [10].

Optimizing Decision Making in Human-centered Systems

Machine Teaching for Human Learners In addition to optimizing the computational and label cost, another important aspect of IML, when considering the interaction with a *human* agent, is the *interpretability* of the model and the *explainability* of the interaction itself. As our first attempt to address interpretability, we started to investigate machine teaching, with the goal of optimally interacting with a (human) learner. We have developed novel machine teaching algorithms for a variety of teaching scenarios, including interpretable teaching with explanations [4, 2], adaptive teaching for forgetful learners [15], and novel interactive models for algorithmic machine teaching [11, 1]. In collaboration with the computer vision lab at Caltech, we evaluated our algorithms on real human subjects and demonstrated promising results in teaching non-Chinese speaker recognizing Chinese characters [2]. In an ongoing collaboration with neurobiologists at Caltech, we are investigating automated teaching algorithms for training mice performing maze tasks.

Future Directions: Learning and Decision Making in Complex Environments

As discussed in my prior work, I am interested in developing interactive, intelligent decision making systems that are motivated by *real-world applications*, yield *new theoretical insights*, and demonstrate *tangible practical impacts*. While the existing results are encouraging, they also raise open questions. In the following, I outline a few promising problems I intend to pursue in this direction.

Optimizing Decision Making with Rich Interfaces

Many real-world applications are structured as they are composed of multiple correlated random variables. For example, in computer vision, we might want to predict the semantic category of each pixel; in natural language processing, we might be interested in parsing sentences syntactically. Traditional learning algorithms for structured problems tackle the supervised setting, where input-output pairs are given and each structured output is fully labeled [17]. However, due to its intrinsic computational difficulties as we have to deal with exponentially sized output spaces, little has been known about such problems under the (inter-) active learning setting.

Moreover, *explanatory* labels often carry implicit structural information about the data points; hence incorporating the explanatory feedback may dramatically reduce the uncertainty of the output space, making it feasible to analyze active learning for such complex hypotheses. Consider a scene understanding task in computer vision, where a learner queries the label of an image segment. Rather than return the partial label "*tree*", the domain expert may further provide support sentence s: "*because it appears above the ground and below the sky*". A key challenge of including s is that the complexity of the problem increases by an order of magnitude. I will investigate new algorithms for exploiting such structural information of explanatory feedback.

Learning to Make Decisions

When the optimization cost of making the best decisions is prohibitive, it is desirable to design a learning system that can learn from previous decision histories. *Learning from demonstration* (LfD), or imitation learning, explores techniques for learning a controller (i.e., a mapping from states to actions) *directly* from examples provided by a human instructor [12]. It has been extensively studied for structured prediction problems. There are three types of queries for LfD [3]: *label queries* (as in *active instance labeling*, e.g., "what is the label of this data point?"), *demonstration queries* (as in *active class selection* [19], e.g., "can you show me an example of class A?"), and *feature queries* (as in *active feature evaluation* [13], e.g., "is this feature important for the target concept being learned?"). Since the instructor plays a powerful role in the learning process and the cost of interacting with the instructor is usually high, it is important for the learner to realize which are the "informative" queries at different stages of the learning process. However, few LfD algorithms have been utilized in an iterative setting. Understanding how to ask questions in a principled manner, offers some of the most promising areas of future work for the design of LfD systems. I will consider queries of mixed-granularity and develop active learning algorithms under the LfD paradigm.

Collaborative Machine and Human Learning

Interleaving Active Learning with Machine Teaching In interactive, interpretable machine learning systems, learning and teaching are reciprocal and tightly coupled. On the one hand, the system is required to intelligently issue queries that are *informative* for learning complex hypotheses; on the other hand, the learner ought to be clear about what it means by the proposed queries, by providing additional explanations that ensure interpretability. This motivates me to weave together techniques developed for both active learning (i.e., in proposing the query) and machine teaching (i.e., in generating the explanation of a query). An interesting future work is to develop a macro framework that jointly optimizes *the label complexity* and *the model interpretability*. It also provides the opportunity to explore structured machine teaching settings that unify prior work on structured prediction and machine teaching, which has thus far focused on extremely simple concept classes.

Learning to Explain when Teaching Human Learners We have demonstrated that *explanations* could be used as a powerful tool for communicating with human learners in machine teaching tasks [4]. While an optimization algorithm is capable of computing a (near-) optimal sequence of training examples for specifying a concept/hypothesis, generating interpretable explanations, i.e., about why a label is assigned to an example, remains a challenging task. Human learners, on the contrary, although subject to physical constraints (such as limitation in memory), are good at providing explanations even during the learning process. In a classroom setting, it is appealing to consider a collaborative learning system that can jointly *learn to generate explanations* from a population of human learners, while *optimizing the teaching performance* for the population. More generally, I am interested in building collaborative learning approaches for machine teaching which accommodates the learners' cognitive models and rich interfaces.

References

[1] M. Ahmadi, B. Wu, Y. Chen, Y. Yue, and U. Topcu. Barrier certificates for assured machine teaching. *CoRR*, abs/1810.00093, 2018.

[2] O. M. Aodha, S. Su, Y. Chen, P. Perona, and Y. Yue. Teaching categories to human learners with visual explanations. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, April 2018.

[3] M. Cakmak and A. L. Thomaz. Designing robot learners that ask good questions. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, pages 17–24. ACM, 2012.

[4] Y. Chen, O. M. Aodha, S. Su, P. Perona, and Y. Yue. Near-optimal machine teaching via explanatory teaching sets. In *Proc. International Conference on Artificial Intelligence and Statistics (AISTATS)*, April 2018.

[5] Y. Chen, S. H. Hassani, A. Karbasi, and A. Krause. Sequential information maximization: When is greedy near-optimal? In *Proc. International Conference on Learning Theory (COLT)*, July 2015.

[6] Y. Chen, S. H. Hassani, and A. Krause. Near-optimal bayesian active learning with correlated and noisy tests. In *Proc. International Conference on Artificial Intelligence and Statistics (AISTATS)*, April 2017.

[7] Y. Chen, S. Javdani, A. Karbasi, J. A. Bagnell, S. Srinivasa, and A. Krause. Submodular surrogates for value of information. In *Proc. Conference on Artificial Intelligence (AAAI)*, January 2015.

[8] Y. Chen and A. Krause. Near-optimal batch mode active learning and adaptive submodular optimization. In *International Conference on Machine Learning (ICML)*, 2013.

[9] Y. Chen, J.-M. Renders, M. H. Chehreghani, and A. Krause. Efficient online learning for optimizing value of information: Theory and application to interactive troubleshooting. In *Proc. Conference on Uncertainty in AI (UAI)*, August 2017.

[10] Y. Chen, H. Shioi, C. F. Montesinos, L. P. Koh, S. Wich, and A. Krause. Active detection via adaptive submodularity. In *Proc. International Conference on Machine Learning (ICML)*, 2014.

[11] Y. Chen, A. Singla, O. M. Aodha, P. Perona, and Y. Yue. Understanding the role of adaptivity in machine teaching: The case of version space learners. In *Proc. Conference on Neural Information Processing Systems (NIPS)*, December 2018.

[12] S. Chernova and A. L. Thomaz. Robot learning from human teachers. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 8(3):1–121, 2014.

[13] G. Druck, B. Settles, and A. McCallum. Active learning by labeling features. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1*, pages 81–90, 2009.

[14] D. Golovin and A. Krause. Adaptive submodularity: Theory and applications in active learning and stochastic optimization. *Journal of Artificial Intelligence Research (JAIR)*, 42:427–486, 2011.

[15] A. Hunziker, Y. Chen, O. M. Aodha, M. G. Rodriguez, A. Krause, P. Perona, Y. Yue, and A. Singla. Teaching multiple concepts to forgetful learners. *CoRR*, abs/1805.08322, 2018.

[16] S. Javdani, Y. Chen, A. Karbasi, A. Krause, D. Bagnell, and S. Srinivasa. Near optimal bayesian active learning for decision making. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2014.

[17] T. Joachims, T. Hofmann, Y. Yue, and C.-N. Yu. Predicting structured objects with support vector machines. *Communications of the ACM*, 52(11):97–104, 2009.

[18] D. V. Lindley. On a measure of the information provided by an experiment. *The Annals of Mathematical Statistics*, 1956.

[19] R. Lomasky, C. E. Brodley, M. Aernecke, D. Walt, and M. Friedl. Active class selection. In *European Conference on Machine Learning*, pages 640–647. Springer, 2007.

[20] J. Song, Y. Chen, and Y. Yue. A general framework for multi-fidelity bayesian optimization with gaussian processes. In *Proc. International Conference on Artificial Intelligence and Statistics (AISTATS)*, April 2019.

[21] K. Yang, Y. Chen, A. Lee, and Y. Yue. Batched stochastic bayesian optimization via combinatorial constraints design. In *Proc. International Conference on Artificial Intelligence and Statistics (AISTATS)*, April 2019.