DDA6201 Online Algorithms and Online Learning
Course Project Topics

Basic Scheme (please read carefully)

- **Grading policy:** The project will be evaluated based on the final report **(40%)** and in-class group presentation **(60%)**. The project has 50 points in total. The report should be written as a technical paper. The length of the report should be **8-10 A4 pages (10pt, single column, normal spacing, excluding references and appendices)**.

  Every presentation should be finished in 30 minutes. There will be a 5-minute QA session after each presentation.

  The evaluation of the project will be based on the following four aspects: **significance, novelty, technical soundness, completeness**.

- **Submission Format:**
  - **Report Templates:** `https://www.overleaf.com/latex/templates/neurips-2023/vstgtvjwgdng`
  - The file name of your presentation slides/PPT/PDF should be in the form of '**Pre_FirstN_LastN**'.
  - The file name of your report PDF/zip should be in the form of '**Report_FirstN_LastN**'.
  - Submit a zip file '**Project_FirstN_LastN**' with the slides, report, and your source code if there is any to `litongxin@cuhk.edu.cn`.

- **Evaluation criteria (for both report and presentation) and guidelines:**
  - **significance (25%):** the problem you studied should be interesting and hasn't been well-solved.
  - **novelty (25%):** at least one aspect of your problem, data, and method is novel. Do not apply an existing method to an existing dataset to solve an existing problem. You may collect or create a new dataset, propose a new problem, devise a new method, or modify an existing method.
  - **technical soundness (25%):** the techniques of data collection, processing, modeling, analyzing, etc., you used, should be reasonable, correct, and explainable.
  - **completeness (25%):** your report and presentation should contain the necessary and important information, description, explanation, or/and discussion, etc., about your task, data, method, and results.

- **Deadlines:** Detailed presentation schedules and report submission deadlines will be planned and announced.

Below we exemplify some potential project topics. **Students are free to choose any topics related to online algorithms and online learning, not limited to the following examples.**

1. **RLHF theory**

   Reinforcement Learning from Human Feedback (RLHF) [12, 8] represents a significant shift in how reinforcement learning (RL) systems are trained, focusing on leveraging human feedback to guide the learning process in environments where traditional reward signals may be sparse or misaligned with desired outcomes. By comparing RLHF's regret metrics with those from conventional RL approaches, this topic aims to uncover deeper theoretical insights and practical implications, enhancing our understanding of RLHF's efficacy and efficiency. Treating the training of an RL agent with human feedback as an online process, it would be interesting to analyze the robustness of RLHF compared to standard RL models. The work in [24] proves for a wide range of preference models, it is possible to solve preference-based RL directly using existing algorithms and techniques for reward-based RL, with small or no extra costs. See the recent results in [26, 23]. A simpler problem is the dueling bandits problem [27, 22], which we will probably cover in our future lectures.

2. **Online algorithms with predictions**

   The concept of integrating black-box machine-learned guidance into online algorithms was initially introduced by [18]. [20] coined terms "robustness" and "consistency" with formal mathematical definitions based on the competitive ratio. Over the past few years, the consistency and robustness approach has gained widespread popularity and has been utilized to design online algorithms with black-box advice for various applications, including ski rental [20, 25, 5],caching [21, 17, 13], bipartite matching [3], online covering [4, 2], and convex body chasing [9]. etc. See [19] and the website below for more details. `https://algorithms-with-predictions.github.io/` Here, we list a few interesting directions to explore:

   - Learning predictions for algorithms with predictions [14]
   - Algorithms with multiple predictions [2]
   - Algorithms with distributional advice [10]
   - Value function predictions in reinforcement learning [11, 15]

   Interesting future directions include understanding the impact of model complexity. Besides, it'd be interesting to see how predictions from LLMs and public model can be augmented in downstream decision-making tasks, with privacy/fairness guarantees.

3. **Beyond consistency and robustness tradeoffs**

   The limitations of current consistency and robustness metrics in online learning problems as outlined in [16], especially in the context of control and reinforcement learning (RL) models, highlights a critical gap in the evaluation of these systems. In scenarios where prediction errors can be (explicitly or implicitly) measured, traditional metrics (see Figure 1) may fail to provide a comprehensive assessment of a model's performance. This project aims to develop a generalized framework for characterizing the trade-offs between consistency (the model's performance following predictions closely) and robustness (the model's resilience to prediction errors) in a way that is more applicable to control and RL models when the prediction error is observable.

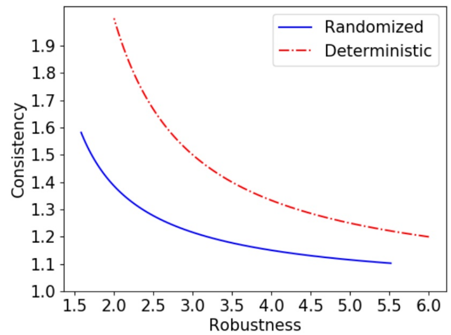

Figure 1: Traditional robustness and consistency tradeoff for the ski-rental problem in [20].

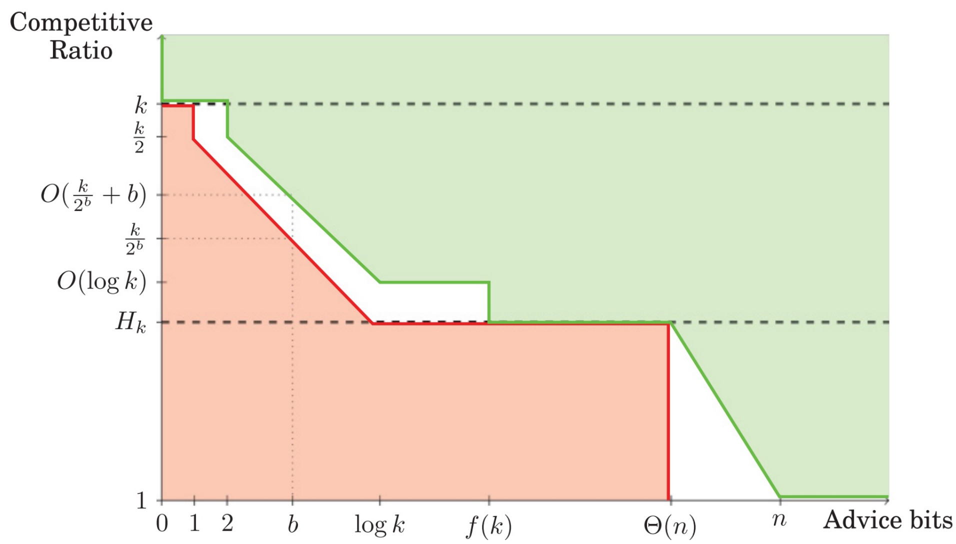

Figure 2: The (asymptotic) tradeoff between competitive ratio and advice for PAGING in [6].

4. **Algorithm with predictions and advice**

The concept of quantifying predictions or advice in online algorithms in terms of bits, as explored by Boyar et al. [6], offers a fascinating intersection between information theory and algorithm design. This research aims to deepen the understanding of how predictions and advice can be quantified and utilized within online algorithms, particularly focusing on the impact of this quantification on algorithmic performance and decision-making processes. By leveraging information-theoretic measures, this project seeks to characterize the utility and efficiency of predictions/advice and the impact on consistency and robustness tradeoffs based on quantified information.

5. **Online Neural Network Design**

The deployment of neural networks in real-world decision-making systems, such as power system control or autonomous driving, faces significant challenges due to the static and generalized nature of conventional neural network designs. These systems require models that can dynamically adapt to new, unforeseen conditions, especially when encountering out-of-distribution (OOD) data that was not present during training. This topic aims to explore the development of adaptive online neural network designs that can self-tune in response to evolving data distributions and environmental conditions, thereby achieving optimal performance across both expected and unexpected scenarios. By focusing on adaptive online neural network design, this project aims to bridge the gap between current AI capabilities and the demands of real-world decision-making systems, potentially revolutionizing how neural networks are applied in critical domains.

6. **Applications**

   Data pricing for machine learned real-time flexibility signal. In the context of machine learning and decision support systems, real-time flexibility signals represent a critical type of data that can significantly enhance operational efficiency, particularly in dynamic environments like energy markets, traffic control, and supply chain management. These signals, which indicate the ability to quickly adapt to changes, are valuable for predictive models and real-time decision-making processes. However, establishing effective pricing strategies for such data poses a complex challenge, as it involves considering the timeliness, accuracy, and predictive value of the information. This research project aims to explore innovative data pricing models specifically designed for real-time flexibility signals, taking into account the unique characteristics of these data types and their impact on machine learning outcomes. This research topic is positioned at the intersection of data economics and machine learning, offering the potential to address a significant gap in the efficient use of real-time flexibility signals. By developing and evaluating dynamic pricing strategies for these valuable data types, the project aims to enhance the operational efficiency of machine learning systems across a wide range of applications.

7. **Distributed online decision-making**

   Distributed online decision-making systems are crucial for managing complex, dynamic environments where decisions must be made quickly and efficiently across multiple agents or nodes. These systems find applications in a wide array of fields, including robotics, autonomous vehicle fleets, smart grids, and distributed computing. Integrating Multi-Agent Reinforcement Learning (MARL) (see [28]) with distributed control principles offers a promising avenue for enhancing the autonomy, efficiency, and scalability of these systems. This topic aims to explore novel approaches to distributed online decision-making, focusing on the synergies between MARL and distributed control techniques to address coordination, scalability, and adaptability challenges. By tackling the theoretical and practical challenges associated with these systems, the project aims to pave the way for more autonomous, efficient, and scalable distributed systems across a variety of domains.

8. **Quantum online learning**

   Another interesting topic is the online learning with quantum states. See [1, 7, 29] for more details.

9. **Other online problems related to your research**

References

[1] Scott Aaronson, Xinyi Chen, Elad Hazan, Satyen Kale, and Ashwin Nayak. Online learning of quantum states. *Advances in neural information processing systems*, 31, 2018.

[2] Keerti Anand, Rong Ge, Amit Kumar, and Debmalya Panigrahi. Online algorithms with multiple predictions. In *International Conference on Machine Learning*, pages 582–598. PMLR, 2022.

[3] Antonios Antoniadis, Christian Coester, Marek Elias, Adam Polak, and Bertrand Simon. Online metric algorithms with untrusted predictions. In *International Conference on Machine Learning*, pages 345–355. PMLR, 2020.

[4] Etienne Bamas, Andreas Maggiori, and Ola Svensson. The primal-dual method for learning augmented algorithms. *Advances in Neural Information Processing Systems*, 33:20083–20094, 2020.

[5] Shom Banerjee. Improving online rent-or-buy algorithms with sequential decision making and ml predictions. *Advances in Neural Information Processing Systems*, 33:21072–21080, 2020.

[6] Joan Boyar, Lene M Favrholdt, Christian Kudahl, Kim S Larsen, and Jesper W Mikkelsen. Online algorithms with advice: a survey. *ACM Computing Surveys (CSUR)*, 50(2):1–34, 2017.

[7] Xinyi Chen, Elad Hazan, Tongyang Li, Zhou Lu, Xinzhao Wang, and Rui Yang. Adaptive online learning of quantum states. *arXiv preprint arXiv:2206.00220*, 2022.

[8] Paul F Christiano, Jan Leike, Tom Brown, Miljan Martic, Shane Legg, and Dario Amodei. Deep reinforcement learning from human preferences. *Advances in neural information processing systems*, 30, 2017.

[9] Nicolas Christianson, Tinashe Handina, and Adam Wierman. Chasing convex bodies and functions with black-box advice. In *Conference on Learning Theory*, pages 867–908. PMLR, 2022.

[10] Ilias Diakonikolas, Vasilis Kontonis, Christos Tzamos, Ali Vakilian, and Nikos Zarifis. Learning online algorithms with distributional advice. In *International Conference on Machine Learning*, pages 2687–2696. PMLR, 2021.

[11] Noah Golowich and Ankur Moitra. Can q-learning be improved with advice? In *Conference on Learning Theory*, pages 4548–4619. PMLR, 2022.

[12] Shane Griffith, Kaushik Subramanian, Jonathan Scholz, Charles L Isbell, and Andrea L Thomaz. Policy shaping: Integrating human feedback with reinforcement learning. *Advances in neural information processing systems*, 26, 2013.

[13] Sungjin Im, Ravi Kumar, Aditya Petety, and Manish Purohit. Parsimonious learning-augmented caching. In *International Conference on Machine Learning*, pages 9588–9601. PMLR, 2022.

[14] Misha Khodak, Maria-Florina F Balcan, Ameet Talwalkar, and Sergei Vassilvitskii. Learning predictions for algorithms with predictions. *Advances in Neural Information Processing Systems*, 35:3542–3555, 2022.

[15] Tongxin Li, Yiheng Lin, Shaolei Ren, and Adam Wierman. Beyond black-box advice: Learning-augmented algorithms for mdps with q-value predictions. *Advances in Neural Information Processing Systems*, 36, 2024.

[16] Tongxin Li, Ruixiao Yang, Guannan Qu, Guanya Shi, Chenkai Yu, Adam Wierman, and Steven Low. Robustness and consistency in linear quadratic control with untrusted predictions. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 6(1):1–35, 2022.

[17] Thodoris Lykouris and Sergei Vassilvitskii. Competitive caching with machine learned advice. *Journal of the ACM (JACM)*, 68(4):1–25, 2021.

[18] Mohammad Mahdian, Hamid Nazerzadeh, and Amin Saberi. Online optimization with uncertain information. *ACM Transactions on Algorithms (TALG)*, 8(1):1–29, 2012.

[19] Michael Mitzenmacher and Sergei Vassilvitskii. Algorithms with predictions. *Communications of the ACM*, 65(7):33–35, 2022.

[20] Manish Purohit, Zoya Svitkina, and Ravi Kumar. Improving online algorithms via ml predictions. *Advances in Neural Information Processing Systems*, 31, 2018.

[21] Dhruv Rohatgi. Near-optimal bounds for online caching with machine learned advice. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1834–1845. SIAM, 2020.

[22] Aadirupa Saha, Tomer Koren, and Yishay Mansour. Adversarial dueling bandits. In *International Conference on Machine Learning*, pages 9235–9244. PMLR, 2021.

[23] Chaoqi Wang, Yibo Jiang, Chenghao Yang, Han Liu, and Yuxin Chen. Beyond reverse kl: Generalizing direct preference optimization with diverse divergence constraints. *arXiv preprint arXiv:2309.16240*, 2023.

[24] Yuanhao Wang, Qinghua Liu, and Chi Jin. Is rlhf more difficult than standard rl? a theoretical perspective. *Advances in Neural Information Processing Systems*, 36, 2024.

[25] Alexander Wei and Fred Zhang. Optimal robustness-consistency trade-offs for learning-augmented online algorithms. *Advances in Neural Information Processing Systems*, 33:8042–8053, 2020.

[26] Wei Xiong, Hanze Dong, Chenlu Ye, Han Zhong, Nan Jiang, and Tong Zhang. Gibbs sampling from human feedback: A provable kl-constrained framework for rlhf. *arXiv preprint arXiv:2312.11456*, 2023.

[27] Yisong Yue, Josef Broder, Robert Kleinberg, and Thorsten Joachims. The k-armed dueling bandits problem. *Journal of Computer and System Sciences*, 78(5):1538–1556, 2012.

[28] Kaiqing Zhang, Zhuoran Yang, and Tamer Başar. Multi-agent reinforcement learning: A selective overview of theories and algorithms. *Handbook of reinforcement learning and control*, pages 321–384, 2021.

[29] Julian Zimmert, Naman Agarwal, and Satyen Kale. Pushing the efficiency-regret pareto frontier for online learning of portfolios and quantum states. In *Conference on Learning Theory*, pages 182–226. PMLR, 2022.