Syllabus

Special Topics Seminar: Deep Reinforcement Learning
CS 8803, Spring 2024

Class Sessions
3 Credit Hours
Lecture Hours: 14:00 - 15:15 TTh at Weber III Room 2

Instructor
Instructor: Animesh Garg (office: CODA S1145)
Teaching Assistants:

- Albert Wilcox
- Liquan Wang
- Uzair Akbar

Instructor Office Hours: Tues 330-430, after lecture.
Contact: Use Ed or Canvas to message course staff. *Direct emails may not be answered.*

Logistics
Ed Discussion. This should be your first stop for questions and announcements.
Canvas will be used to take quizzes, view grades, and view assignments.
Course Webpage: https://pairlab.github.io/cs8803-drl-f24

Please refer to the course website for updated lecture schedule and additional policies.

Description: Robots of the future will need to operate autonomously in unstructured and unseen environments. It is imperative that these systems are built on intelligent and adaptive algorithms. Learning by interaction through reinforcement offers a natural mechanism to postulate these problems.

This graduate-level seminar course will cover topics and new research frontiers in reinforcement learning (RL). Planned topics include: Model-Based and Model-Free RL, Policy Search, Monte Carlo Tree Search, off policy evaluation, temporal abstraction/hierarchical approaches, inverse reinforcement learning and imitation learning.

Learning objectives: At the end of this course, you will:
1. Acquire familiarity with state of the art in RL
2. Articulate limitations of current work, identify open frontiers, and scope research projects.
3. Constructively critique research papers, and deliver a tutorial style presentation.
4. Work on a research based project, implement & evaluate experimental results, and discuss future work in a project paper.

Grading & Evaluation:
This course will consist of lectures, along with paper presentations & discussions.
Along with this there would be homeworks, in-class quizzes, and a group project.

Homework(s): 20%

Two homework assignments: programming + short questions
Paper Presentation and implementation: 30%
Quizzes & Participation (live only): 10%
Project: 40%
Proposal (5%)
Intermediate progress (5%)
Presentation (10%)
Final report (20%)

Textbooks & Resources:
There is no official textbook for the class.

A number of the supporting readings will come from:
Reinforcement Learning: An Introduction, Sutton and Barto, 2nd Edition.
This is available for free here and references will refer to the final pdf version available here.

Some other additional references that may be useful are listed below:
1. Reinforcement Learning: State-of-the-Art, Marco Wiering & Martijn van Otterlo, Eds. [link]
2. Artificial Intelligence: A Modern Approach, Stuart J. Russell and Peter Norvig.[link]
3. Deep Learning, Ian Goodfellow, Yoshua Bengio, and Aaron Courville. [link]

Additional Resources from similar courses.
1. RL Course from UofT - Animesh Garg [link]
2. RL Course from UW - Byron Boots [link]
3. RL Course from Stanford - Emma Brunskill [link]
4. RL Course from University of Alberta - Martha White [link]
5. RL course at ASU/MIT, Dimitry Bertsekas -- [Link]
6. David Silver's course on Reinforcement Learning [link]
7. Deep RL Course from Berkeley - Sergey Levine [link]

Prerequisites:

You need to be comfortable with: introductory machine learning concepts (such as from CS7641 or equivalent), linear algebra, basic multivariable calculus, intro to probability. You also need to have strong programming skills in Python. Note: if you don't meet all the prerequisites above please contact the instructor by email. Optional, but recommended: experience with neural networks, such as from CS 4644 / 7643 or equivalent, introductory-level familiarity with reinforcement learning and control.

Academic Integrity
Academic dishonesty will not be tolerated. This includes cheating, lying about course matters, plagiarism, or helping others commit a violation of the Honor Code. Plagiarism includes reproducing the words of others without both the use of quotation marks and citations. Students are reminded of the obligations and expectations associated with the Georgia Tech Academic Honor Code and Student Code of Conduct. For exams, no supporting material is allowed, unless specified, and no electronic materials are allowed (phones, laptops, tablets, etc). A stand-alone calculator may be used.

You are expected to implement the core components of each project on your own, but the extra credit opportunities may build on third party data sets or code. That's acceptable. Feel free to include results built on other software, as long as what you hand in clearly cites the third-party source, making it clear it is not your own work.

You should not view or edit anyone else's code. You should not post code to Ed Discussion, except for starter code / helper code that isn't related to the core project.

Learning Accommodations
If needed, we will make accommodations for students with documented disabilities. These accommodations must be arranged in advance and in accordance with the Office of Disability Services policies.

Suggested Paper List

1. Policy Gradient Methods

- Paper #1: Schulman, J., Moritz, P., Levine, S., Jordan, M. and Abbeel, P., 2015. High-dimensional continuous control using generalized advantage estimation. arXiv preprint arXiv:1506.02438.
- Paper #2: Schulman, J., Wolski, F., Dhariwal, P., Radford, A. and Klimov, O., 2017. Proximal policy optimization algorithms. arXiv preprint arXiv:1707.06347.
- Paper #3: Mania, Horia, Aurelia Guy, and Benjamin Recht. "Simple random search provides a competitive approach to reinforcement learning." arXiv preprint arXiv:1803.07055 (2018).
- Runner-up: Schulman, J., Levine, S., Abbeel, P., Jordan, M. and Moritz, P., 2015, June. Trust region policy optimization. In International conference on machine learning (pp. 1889-1897). PMLR.
- Runner-up: Williams, R.J., 1992. Simple statistical gradient-following algorithms for connectionist reinforcement learning. Machine learning, 8(3), pp.229-256.

2. Value-function Methods

- Paper #1: Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., Graves, A., Riedmiller, M., Fidjeland, A.K., Ostrovski, G. and Petersen, S., 2015. Human-level control through deep reinforcement learning. nature, 518(7540), pp.529-533.
- Paper #2: Lillicrap, T.P., Hunt, J.J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D. and Wierstra, D., 2015. Continuous control with deep reinforcement learning. arXiv preprint arXiv:1509.02971.
- Paper #3: Haarnoja, T., Zhou, A., Hartikainen, K., Tucker, G., Ha, S., Tan, J., Kumar, V., Zhu, H., Gupta, A., Abbeel, P. and Levine, S., 2018. Soft actor-critic algorithms and applications. arXiv preprint arXiv:1812.05905.
- Runner-up: Hessel, M., Modayil, J., Van Hasselt, H., Schaul, T., Ostrovski, G., Dabney, W., Horgan, D., Piot, B., Azar, M. and Silver, D., 2018, April. Rainbow: Combining improvements in deep reinforcement learning. In Thirty-second AAAI conference on artificial intelligence.
- Runner-up: Fujimoto, S., Hoof, H. and Meger, D., 2018, July. Addressing function approximation error in actor-critic methods. In International Conference on Machine Learning (pp. 1587-1596). PMLR.
- Runner-up: Mnih, V., Badia, A.P., Mirza, M., Graves, A., Lillicrap, T., Harley, T., Silver, D. and Kavukcuoglu, K., 2016, June. Asynchronous methods for deep reinforcement learning. In International conference on machine learning (pp. 1928-1937). PMLR.
- Runner-up: Peng, X.B., Kumar, A., Zhang, G. and Levine, S., 2019. Advantage-weighted regression: Simple and scalable off-policy reinforcement learning. arXiv preprint arXiv:1910.00177.
- Runner-up: Chen, X., Wang, C., Zhou, Z. and Ross, K., 2021. Randomized ensembled double q-learning: Learning fast without a model. arXiv preprint arXiv:2101.05982.

3. Other Techniques (Experience/Curriculum/Architecture)

- Paper #1: Andrychowicz, M., Wolski, F., Ray, A., Schneider, J., Fong, R., Welinder, P., McGrew, B., Tobin, J., Abbeel, P. and Zaremba, W., 2017. Hindsight experience replay. arXiv preprint arXiv:1707.01495.
- Paper #2: Lynch, C., Khansari, M., Xiao, T., Kumar, V., Tompson, J., Levine, S. and Sermanet, P., 2020, May. Learning latent plans from play. In Conference on Robot Learning (pp. 1113-1132). PMLR.

- Paper #3: Wang, R., Lehman, J., Clune, J. and Stanley, K.O., 2019. Paired open-ended trailblazer (poet): Endlessly generating increasingly complex and diverse learning environments and their solutions. arXiv preprint arXiv:1901.01753.
- Runner-up: Xie, Z., Ling, H.Y., Kim, N.H. and van de Panne, M., 2020, December. ALLSTEPS: Curriculum‐driven Learning of Stepping Stone Skills. In Computer Graphics Forum (Vol. 39, No. 8, pp. 213-224).
- Runner-up: Iscen, A., Caluwaerts, K., Tan, J., Zhang, T., Coumans, E., Sindhwani, V. and Vanhoucke, V., 2018, October. Policies modulating trajectory generators. In Conference on Robot Learning (pp. 916-926). PMLR.
- Runner-up: Gaier, A. and Ha, D., 2019. Weight agnostic neural networks. arXiv preprint arXiv:1906.04358.
- Runner-up: Eysenbach, B., Gupta, A., Ibarz, J. and Levine, S., 2018. Diversity is all you need: Learning skills without a reward function. arXiv preprint arXiv:1802.06070.
- Runner-up: Matiisen, T., Oliver, A., Cohen, T. and Schulman, J., 2019. Teacher–student curriculum learning. IEEE transactions on neural networks and learning systems, 31(9), pp.3732-3740.
- Runner-up: Baker, B., Kanitscheider, I., Markov, T., Wu, Y., Powell, G., McGrew, B. and Mordatch, I., 2019. Emergent tool use from multi-agent autocurricula. arXiv preprint arXiv:1909.07528.

4. Meta learning and adaptation

- Paper #1: Duan, Y., Schulman, J., Chen, X., Bartlett, P.L., Sutskever, I. and Abbeel, P., 2016. Rl $^ 2$: Fast reinforcement learning via slow reinforcement learning. arXiv preprint arXiv:1611.02779.
- Paper #2: Finn, C., Abbeel, P. and Levine, S., 2017, July. Model-agnostic meta-learning for fast adaptation of deep networks. In International Conference on Machine Learning (pp. 1126-1135). PMLR.
- Paper #3: Kumar, Ashish, Zipeng Fu, Deepak Pathak, and Jitendra Malik. "Rma: Rapid motor adaptation for legged robots." arXiv preprint arXiv:2107.04034 (2021).
- Runner-up: Yu, W., Tan, J., Liu, C.K. and Turk, G., 2017. Preparing for the unknown: Learning a universal policy with online system identification. arXiv preprint arXiv:1702.02453.
- Runner-up: Yu, W., Tan, J., Bai, Y., Coumans, E. and Ha, S., 2020. Learning fast adaptation with meta strategy optimization. IEEE Robotics and Automation Letters, 5(2), pp.2950-2957.
- Runner-up: Yin, M., Tucker, G., Zhou, M., Levine, S. and Finn, C., 2019. Meta-learning without memorization. arXiv preprint arXiv:1912.03820.

5. Imitation Learning

- Paper #1: Ross, S., Gordon, G. and Bagnell, D., 2011, June. A reduction of imitation learning and structured prediction to no-regret online learning. In Proceedings of the fourteenth international conference on artificial intelligence and statistics (pp. 627-635). JMLR Workshop and Conference Proceedings.
- Paper #2: Chen, D., Zhou, B., Koltun, V. and Krähenbühl, P., 2020, May. Learning by cheating. In Conference on Robot Learning (pp. 66-75). PMLR.
- Paper #3: Merel, J., Tassa, Y., TB, D., Srinivasan, S., Lemmon, J., Wang, Z., Wayne, G. and Heess, N., 2017. Learning human behaviors from motion capture by adversarial imitation. arXiv preprint arXiv:1707.02201.
- Runner-up: Peng, X.B., Abbeel, P., Levine, S. and van de Panne, M., 2018. Deepmimic: Example-guided deep reinforcement learning of physics-based character skills. ACM Transactions on Graphics (TOG), 37(4), pp.1-14.
- Runner-up: Abbeel, P. and Ng, A.Y., 2004, July. Apprenticeship learning via inverse reinforcement learning. In Proceedings of the twenty-first international conference on Machine learning (p. 1).
- Runner-up: Berseth, G., Xie, C., Cernek, P. and Van de Panne, M., 2018.

Progressive reinforcement learning with distillation for multi-skilled motion control. arXiv preprint arXiv:1802.04765.

6. Model-based Reinforcement Learning

- Paper #1: Janner, M., Fu, J., Zhang, M. and Levine, S., 2019. When to trust your model: Model-based policy optimization. arXiv preprint arXiv:1906.08253.
- Paper #2: Hafner, D., Lillicrap, T., Fischer, I., Villegas, R., Ha, D., Lee, H. and Davidson, J., 2019, May. Learning latent dynamics for planning from pixels. In International Conference on Machine Learning (pp. 2555-2565). PMLR.
- Paper #3: Kaiser, L., Babaeizadeh, M., Milos, P., Osinski, B., Campbell, R.H., Czechowski, K., Erhan, D., Finn, C., Kozakowski, P., Levine, S. and Mohiuddin, A., 2019. Model-based reinforcement learning for atari. arXiv preprint arXiv:1903.00374.
- Runner-up: Chua, K., Calandra, R., McAllister, R. and Levine, S., 2018. Deep reinforcement learning in a handful of trials using probabilistic dynamics models. arXiv preprint arXiv:1805.12114.
- Runner-up: Buckman, J., Hafner, D., Tucker, G., Brevdo, E. and Lee, H., 2018. Sample-efficient reinforcement learning with stochastic ensemble value expansion. arXiv preprint arXiv:1807.01675.

7. Sim-to-real Techniques

- Paper #1: Xue Bin Peng, Marcin Andrychowicz, Wojciech Zaremba, Pieter Abbeel, 2018. Sim-to-Real Transfer of Robotic Control with Dynamics Randomization. International Conference on Robotics and Automation (ICRA) 2018.
- Paper #2: Yevgen Chebotar, Ankur Handa, Viktor Makoviychuk, Miles Macklin, Jan Issac, Nathan Ratliff, Dieter Fox, 2019. Closing the Sim-to-Real Loop: Adapting Simulation Randomization with Real World Experience. International Conference on Robotics and Automation (ICRA) 2019.
- Paper #3: Hwangbo, J., Lee, J., Dosovitskiy, A., Bellicoso, D., Tsounis, V., Koltun, V. and Hutter, M., 2019. Learning agile and dynamic motor skills for legged robots. Science Robotics, 4(26), p.eaau5872.
- Runner-up: Wenhao Yu, Jie Tan, C. Karen Liu, Greg Turk, 2017. Preparing for the Unknown: Learning a Universal Policy with Online System Identification. Robotics: Science and Systems (RSS) 2017.
- Runner-up: Konstantinos Bousmalis, Alex Irpan, Paul Wohlhart, Yunfei Bai, Matthew Kelcey, Mrinal Kalakrishnan, Laura Downs, Julian Ibarz, Peter Pastor, Kurt Konolige, Sergey Levine, Vincent Vanhoucke, Using Simulation and Domain Adaptation to Improve Efficiency of Deep Robotic Grasping. IEEE International Conference on Robotics and Automation (ICRA), 2018
- Runner-up: Jie Tan, Tingnan Zhang, Erwin Coumans, Atil Iscen, Yunfei Bai, Danijar Hafner, Steven Bohez, Vincent Vanhoucke. 2018. Sim-to-Real: Learning Agile Locomotion For Quadruped Robots. Robotics: Science and Systems (RSS) 2018.

8. Safety-aware Reinforcement Learning

- Paper #1: Joshua Achiam, David Held, Aviv Tamar, Pieter Abbeel. 2017. Constrained Policy Optimization. Proceedings of the 34th International Conference on Machine Learning (ICML), 2017.
- Paper #2: Brijen Thananjeyan, Ashwin Balakrishna, Suraj Nair, Michael Luo, Krishnan Srinivasan, Minho Hwang, Joseph E. Gonzalez, Julian Ibarz, Chelsea Finn, Ken Goldberg. 2020. Recovery RL: Safe Reinforcement Learning with Learned Recovery Zones. Robotics and Automation Letters (RA-L) Journal and International Conference on Robotics and Automation (ICRA), 2021
- Paper #3: Homanga Bharadhwaj, Aviral Kumar, Nicholas Rhinehart, Sergey Levine, Florian Shkurti, Animesh Garg, 2021. Conservative Safety Critics for Exploration. ICLR 2021.
- runner-up: Dalal, G., Dvijotham, K., Vecerík, M., Hester, T., Paduraru, C., & Tassa, Y.

(2018). Safe Exploration in Continuous Action Spaces. ArXiv, abs/1801.08757.

9. Computer Animation:

- Paper #1: Liu, L. and Hodgins, J., 2017. Learning to schedule control fragments for physics-based characters using deep q-learning. ACM Transactions on Graphics (TOG), 36(3), pp.1-14.
- Paper #2: Won, J., Gopinath, D. and Hodgins, J., 2020. A scalable approach to control diverse behaviors for physically simulated characters. ACM Transactions on Graphics (TOG), 39(4), pp.33-1.
- Paper #3: Peng, X.B., Berseth, G., Yin, K. and Van De Panne, M., 2017. Deeploco: Dynamic locomotion skills using hierarchical deep reinforcement learning. ACM Transactions on Graphics (TOG), 36(4), pp.1-13.

10. Foundational Models

- Paper #1: Ahn, M., Brohan, A., Brown, N., Chebotar, Y., Cortes, O., David, B., Finn, C., Fu, C., Gopalakrishnan, K., Hausman, K. and Herzog, A., 2022. Do as i can, not as i say: Grounding language in robotic affordances. arXiv preprint arXiv:2204.01691.
- Paper #2: Brohan, A., Brown, N., Carbajal, J., Chebotar, Y., Chen, X., Choromanski, K., Ding, T., Driess, D., Dubey, A., Finn, C. and Florence, P., 2023. Rt-2: Vision-language-action models transfer web knowledge to robotic control. arXiv preprint arXiv:2307.15818.
- Paper #3: Chi, C., Feng, S., Du, Y., Xu, Z., Cousineau, E., Burchfiel, B. and Song, S., 2023. Diffusion policy: Visuomotor policy learning via action diffusion. arXiv preprint arXiv:2303.04137.

Other Papers

- -- locomotion --
- Paper #1: Lee, J., Hwangbo, J., Wellhausen, L., Koltun, V. and Hutter, M., 2020. Learning quadrupedal locomotion over challenging terrain. Science robotics, 5(47).
- Paper #2: Siekmann, J., Green, K., Warila, J., Fern, A. and Hurst, J., 2021. Blind Bipedal Stair Traversal via Sim-to-Real Reinforcement Learning. arXiv preprint arXiv:2105.08328.
- Runner-up: Yu, W., Jain, D., Escontrela, A., Iscen, A., Xu, P., Coumans, E., Ha, S., Tan, J. and Zhang, T., 2021, June. Visual-Locomotion: Learning to Walk on Complex Terrains with Vision. In 5th Annual Conference on Robot Learning.
- Runner-up: Xie, Z., Da, X., Babich, B., Garg, A. and van de Panne, M., 2021. GLiDE: Generalizable Quadrupedal Locomotion in Diverse Environments with a Centroidal Model. arXiv preprint arXiv:2104.09771.
- -- manipulation --
- Paper #1: Kalashnikov, D., Irpan, A., Pastor, P., Ibarz, J., Herzog, A., Jang, E., Quillen, D., Holly, E., Kalakrishnan, M., Vanhoucke, V. and Levine, S., 2018. Qt-opt: Scalable deep reinforcement learning for vision-based robotic manipulation. arXiv preprint arXiv:1806.10293.
- Paper #2: Lee, M.A., Zhu, Y., Srinivasan, K., Shah, P., Savarese, S., Fei-Fei, L., Garg, A. and Bohg, J., 2019, May. Making sense of vision and touch: Self-supervised learning of multimodal representations for contact-rich tasks. In 2019 International Conference on Robotics and Automation (ICRA) (pp. 8943-8950). IEEE.
- Runner-up: Xia, F., Li, C., Martín-Martín, R., Litany, O., Toshev, A. and Savarese, S., 2020. Relmogen: Leveraging motion generation in reinforcement learning for mobile manipulation. arXiv preprint arXiv:2008.07792.
- -- Navigation --
- Paper #1: Mirowski, P., Pascanu, R., Viola, F., Soyer, H., Ballard, A.J., Banino, A., Denil, M., Goroshin, R., Sifre, L., Kavukcuoglu, K. and Kumaran, D., 2016. Learning

to navigate in complex environments. arXiv preprint arXiv:1611.03673.

- Paper #2: Wijmans, E., Kadian, A., Morcos, A., Lee, S., Essa, I., Parikh, D., Savva, M. and Batra, D., 2019. Dd-ppo: Learning near-perfect pointgoal navigators from 2.5 billion frames. arXiv preprint arXiv:1911.00357.
- Runner-up: Sorokin, M., Tan, J., Liu, C.K. and Ha, S., 2021. Learning to Navigate Sidewalks in Outdoor Environments. arXiv preprint arXiv:2109.05603.
-