CS394R
Reinforcement Learning: Theory and Practice

Scott Niekum and Peter Stone

Department of Computer Science
The University of Texas at Austin

Good Morning Colleagues

- Are there any questions?

Logistics

- Last step: final projects!

Logistics

- Last step: final projects!

  – Build on proposal and literature survey

Logistics

- Last step: final projects!

  - Build on proposal and literature survey

- This week:

  - Evaluation and reproducibility

Logistics

- Last step: final projects!

    – Build on proposal and literature survey

- This week:

    – Evaluation and reproducibility
    – Review and big picture

Logistics

- Last step: final projects!

  - Build on proposal and literature survey

- This week:

  - Evaluation and reproducibility
  - Review and big picture

- Do the course evaluation surveys!

Discussion Questions (for Thursday)

- Does RL matter?

Discussion Questions (for Thursday)

- Does RL matter?

- What were your biggest misconceptions about RL?

Discussion Questions (for Thursday)

- Does RL matter?

- What were your biggest misconceptions about RL?

- Are there any important ethical concerns related to RL?

Discussion Questions (for Thursday)

- Does RL matter?

- What were your biggest misconceptions about RL?

- Are there any important ethical concerns related to RL?

- And others...

Common Questions

- How big of an issue is overclaiming in the RL (and broader ML) community? Is it better to overclaim or underclaim?

Common Questions

- How big of an issue is overclaiming in the RL (and broader ML) community? Is it better to overclaim or underclaim?

    "A barrier to adoption of evaluation protocols proposed in this work, and more generally, rigorous evaluation, is whether there are clear incentives for researchers to do so, as more rigor generally entails more nuanced and tempered claims."
    (Agarwal et al.)

Common Questions

- How big of an issue is overclaiming in the RL (and broader ML) community? Is it better to overclaim or underclaim?

  "A barrier to adoption of evaluation protocols proposed in this work, and more generally, rigorous evaluation, is whether there are clear incentives for researchers to do so, as more rigor generally entails more nuanced and tempered claims."
  (Agarwal et al.)

- Can we consider papers whose codebases aren't published trustworthy?

What's the Goal?

- Is there a set of environments in RL that are sufficient to show good performance of a method?

What's the Goal?

- Is there a set of environments in RL that are sufficient to show good performance of a method?

- How much should we care about learning methods that can perform well only with very particular hyperparameters? Is some resilience to hyperparameter change important for robust RL?

What's the Goal?

- Is there a set of environments in RL that are sufficient to show good performance of a method?

- How much should we care about learning methods that can perform well only with very particular hyperparameters? Is some resilience to hyperparameter change important for robust RL?

Some details

- How practical it is to get really good statistical results using few runs?

- Please discuss the statistical methods mentioned in the papers in greater detail

Some details

- How practical it is to get really good statistical results using few runs?

- Please discuss the statistical methods mentioned in the papers in greater detail

- What role does randomness play in reproducibility? Should a random seeds be reported? Are stochasticity and reproducibility really at odds with each other?

Some details

- How practical it is to get really good statistical results using few runs?

- Please discuss the statistical methods mentioned in the papers in greater detail

- What role does randomness play in reproducibility? Should a random seeds be reported? Are stochasticity and reproducibility really at odds with each other?

  – **Reproducibility:** independent researchers arriving at the same results using their own data and methods

Some details

- How practical it is to get really good statistical results using few runs?

- Please discuss the statistical methods mentioned in the papers in greater detail

- What role does randomness play in reproducibility? Should a random seeds be reported? Are stochasticity and reproducibility really at odds with each other?

  - **Reproducibility:** independent researchers arriving at the same results using their own data and methods
  - **Replicability:** a different team arriving at the same results using the original author's artifacts

Deep RL that Matters

- Policy Gradient: TRPO, PPO, DDPG, ACKTR
    - MLPs, default params

Deep RL that Matters

- Policy Gradient: TRPO, PPO, DDPG, ACKTR

  – MLPs, default params

- Continuos Domains: Hopper and Half-Cheetah

Hyperparameters

- Intrinsic hyperparameters:

Hyperparameters

- Intrinsic hyperparameters:

  - Network architecture — huge variations

Hyperparameters

- Intrinsic hyperparameters:

  – Network architecture — huge variations
  – Reward rescaling can have a large effect, interact with other hyperparams

Hyperparameters

- Intrinsic hyperparameters:

  – Network architecture — huge variations
  – Reward rescaling can have a large effect, interact with other hyperparams
  – Random seeds and environment matter!

Hyperparameters

- Intrinsic hyperparameters:

  - Network architecture — huge variations
  - Reward rescaling can have a large effect, interact with other hyperparams
  - Random seeds and environment matter!

- Extrinsic hyperparameters:

  - Hyperparameters, codebase

Hyperparameters

- Intrinsic hyperparameters:

    – Network architecture — huge variations
    – Reward rescaling can have a large effect, interact with other hyperparams
    – Random seeds and environment matter!

- Extrinsic hyperparameters:

    – Hyperparameters, codebase

- Metrics:

    – Online vs. offline evaluation
    – Bootstrapping for confidence intervals, significance testing

Recommendations

- Try to find hyperparams that match original performance (be fair)

Recommendations

- Try to find hyperparams that match original performance (be fair)

- Use many random seeds, test significance

Recommendations

- Try to find hyperparams that match original performance (be fair)

- Use many random seeds, test significance

- **report EVERYTHING**

Recommendations

- Try to find hyperparams that match original performance (be fair)

- Use many random seeds, test significance

- **report EVERYTHING**

- Build hyperparameter-agnostic algorithms

Recommendations

- Try to find hyperparams that match original performance (be fair)

- Use many random seeds, test significance

- **report EVERYTHING**

- Build hyperparameter-agnostic algorithms

- Find better evaluation methods

Recommendations

- Try to find hyperparams that match original performance (be fair)

- Use many random seeds, test significance

- **report EVERYTHING**

- Build hyperparameter-agnostic algorithms

- Find better evaluation methods

- Perhaps each algorithm has its place - present algorithms with real-world environments they're good for (e.g. don't try to find 1 universally best alg)

Deep RL at Edge of Statistics

- Environment: Atari 100K benchmark

Deep RL at Edge of Statistics

- Environment: Atari 100K benchmark

- 5 algorithmss: used in Atari

Deep RL at Edge of Statistics

- Environment: Atari 100K benchmark

- 5 algorithmss: used in Atari

- High variability in results

Deep RL at Edge of Statistics

- Environment: Atari 100K benchmark

- 5 algorithmss: used in Atari

- High variability in results

- Bias in medieans

- Lift can't be detected

Deep RL at Edge of Statistics

- Environment: Atari 100K benchmark

- 5 algorithmss: used in Atari

- High variability in results

- Bias in medieans

- Lift can't be detected

- Changes in evaluation protocol matter

Recommendations

- Bootstrapping (different kind) - resample from distribution many times

Recommendations

- Bootstrapping (different kind) - resample from distribution many times

- Performance profiles (percentage above some threshold)

Recommendations

- Bootstrapping (different kind) - resample from distribution many times

- Performance profiles (percentage above some threshold)

- Better summary stats (point estimates) from performance profiles:

  - Interquartile mean (IQM) instead of mean or median

Recommendations

- Bootstrapping (different kind) - resample from distribution many times

- Performance profiles (percentage above some threshold)

- Better summary stats (point estimates) from performance profiles:
  - Interquartile mean (IQM) instead of mean or median
  - Optimality gap

Recommendations

- Bootstrapping (different kind) - resample from distribution many times

- Performance profiles (percentage above some threshold)

- Better summary stats (point estimates) from performance profiles:

  - Interquartile mean (IQM) instead of mean or median
  - Optimality gap
  - Probability of improvement

Other Interesting Questions

- Haroon Mushtaq: I understand the recommendations made to handle reproducibility; however, why is it not enough to establish a "logical proof" of convergence of an algorithm?

Reproducibility

- Hengfa Lu: When results are questionable, why don't reviewers at least point out this problem and authors add more comprehensive experiments to demonstrate the advantages of their methods?

Reproducibility

- Hengfa Lu: When results are questionable, why don't reviewers at least point out this problem and authors add more comprehensive experiments to demonstrate the advantages of their methods?

- Shubhankar Agarwal: Could there be a third party which validates RL algorithms and makes sure results are reproducible and consistent?

Reproducibility

- Hengfa Lu: When results are questionable, why don't reviewers at least point out this problem and authors add more comprehensive experiments to demonstrate the advantages of their methods?

- Shubhankar Agarwal: Could there be a third party which validates RL algorithms and makes sure results are reproducible and consistent?

- Zirui Tang: What is the approximate proportion of papers that can be reproduced well these years?

Incentives

- Shubhankar Agarwal: Getting something working in real-world takes significant effort and engineering. How should we incentivize researchers to take that extra step to make RL algorithms work in real-world?

Limited resources

- Tanvir Mahmud: How practical it is to get good certainty of performance with few runs?

- Jiaxun Cui: What should we do with limited computation?

Limited resources

- Tanvir Mahmud: How practical it is to get good certainty of performance with few runs?

- Jiaxun Cui: What should we do with limited computation?

- Samuel Thomas: For models with a large number of hyperparameters, how do you do a meaningful sensitivity analysis?

Limited resources

- Tanvir Mahmud: How practical it is to get good certainty of performance with few runs?

- Jiaxun Cui: What should we do with limited computation?

- Samuel Thomas: For models with a large number of hyperparameters, how do you do a meaningful sensitivity analysis?

- Steve Han: Should we be measuring time steps or number of episodes? And do these sample efficiency measures account for computational efficiency?

Environments

- Jordi Ramos Chen: Is there a metric that evaluates the different RL environments? For example, what makes ALE such a good and popular benchmark?

Environments

- Jordi Ramos Chen: Is there a metric that evaluates the different RL environments? For example, what makes ALE such a good and popular benchmark?

- Arnav Iyer: Is there a set of environments that represent the space of "environment types"?

- Cevahir Koprulu: Can we categorize environments in RL literature so that we can have an idea about why an approach fails in certain groups but not in others?

- Kyung-Bin Kwon: Can we come up with guidelines for which algorithms work on which types of problems, such as policy gradient vs. DQN?

Paper impacts

- Isha Tarte: Is it common for RL publications to use rliable? Do reviewrs accept point estimates now?

- Nicholas Wolczynski: Has this paper driven meaningful change in the field since 2021? The recommendations the authors give make sense but would add a lot of work and complexity to the papers of the results, is it reasonable to expect such changes?

Hyperparameters

- Matthew Jones: Is it possible to ensure that different RL algorithms are compared to each other using a fair set of hyperparameters for each algorithm?

Hyperparameters

- Matthew Jones: Is it possible to ensure that different RL algorithms are compared to each other using a fair set of hyperparameters for each algorithm?

- Isha Tarte: Are there any hyperparameter agnostic RL algorithms?

Hyperparameters

- Matthew Jones: Is it possible to ensure that different RL algorithms are compared to each other using a fair set of hyperparameters for each algorithm?

- Isha Tarte: Are there any hyperparameter agnostic RL algorithms?

- Srinivas Bangalore Seshadri: Do you think going towards a generalized learning framework (hyperparameter agnostic frameworks) will provide good performance on all problems?

Discussion Points

- Does RL matter?

Discussion Points

- Does RL matter?

- What were your biggest misconceptions about RL?

Discussion Points

- Does RL matter?

- What were your biggest misconceptions about RL?

- Are there any important ethical concerns related to RL?

Discussion Points

- Does RL matter?

- What were your biggest misconceptions about RL?

- Are there any important ethical concerns related to RL?

- What was most difficult about this course?

Discussion Points

- Does RL matter?

- What were your biggest misconceptions about RL?

- Are there any important ethical concerns related to RL?

- What was most difficult about this course?

- What did you like most about it?

Discussion Points

- Does RL matter?

- What were your biggest misconceptions about RL?

- Are there any important ethical concerns related to RL?

- What was most difficult about this course?

- What did you like most about it?

- What would you change?