Research project proposal:

On the sample complexity of Inverse Reinforcement Learning

Filippo Lazzati
filippo.lazzati@asp-poli.it
T2I - Artificial Intelligence

Outline

Introduction to the problem

- **Some general notions**
  - **Artificial Intelligence**
  - **Reinforcement Learning**
  - **Imitation Learning**
  - **Inverse Reinforcement Learning**
  - **Solution Techniques for Inverse Reinforcement Learning**
- Research topic and Problem
  - Research topic
  - Motivations to support the importance of the research topic
  - Description of the problem
  - Motivations to support the importance of the problem

Artificial Intelligence

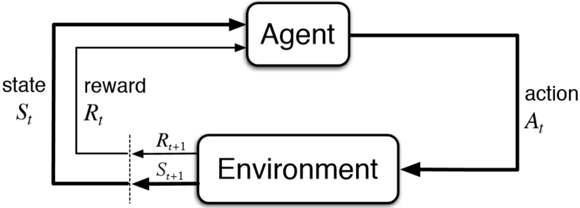
Various approaches (Russell and Norvig 2010)

Think as Humans	Think Rationally
Act as Humans	Act Rationally

Reinforcement Learning

Reinforcement learning *is learning what to do—how to map situations to actions—so as to maximize a numerical reward signal.* (Sutton and Barto 2018)

$$\langle \mathcal{S}, \mathcal{A}, R, p, \mu_0 \rangle \quad \longrightarrow \quad \pi$$



Imitation Learning

Imitation learning is the process of *learning from demonstrations, and the study of algorithms to do so.* (Osa et al. 2018)

- Behavioral Cloning
- Inverse Reinforcement Learning

Inverse Reinforcement Learning

Inverse Reinforcement Learning is *the problem of extracting a reward function given observed, optimal behavior.* (Ng and Russell 2002)

$$\langle \mathcal{S}, \mathcal{A}, p, \mu_0, \pi^E \rangle \quad \longrightarrow \quad R$$

Solution Techniques for Inverse Reinforcement Learning

Margin Optimization maximize the margin between value of observed behavior and the hypothesis

Entropy Optimization maximize the entropy of the distribution over behaviors

Bayesian Update learn posterior over hypothesis space using Bayes rule

Classification and Regression learn a prediction model that imitates observed behavior

Introduction to the problem

- Some general notions
  - Artificial Intelligence
  - Reinforcement Learning
  - Imitation Learning
  - Inverse Reinforcement Learning
  - Solution Techniques for Inverse Reinforcement Learning
- **Research topic and Problem**
  - **Research topic**
  - **Motivations to support the importance of the research topic**
  - **Description of the problem**
  - **Motivations to support the importance of the problem**

Research topic

Sample Complexity means *how much data must we collect in order to achieve "learning"?* (Kakade 2003)

(ϵ, δ)-**correctness** means that the algorithm provides an ϵ-correct solution w.p. at least $1 - \delta$ (Haussler 1990)

Motivations to support the importance of the research topic

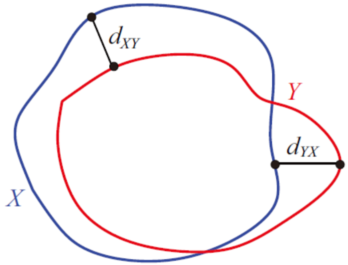
Theory characterize the complexity of the problem

Practice assess the performance of existing and new algorithms

Description of the problem

Feasible set $\mathcal{R}$ is the set of reward functions *compatible with the expert's demonstrations* (Metelli et al. 2021)

(ϵ, δ)-**correctness** after t samples is $\wp\big(h(\mathcal{R}, \hat{\mathcal{R}}_t) \leq \epsilon\big) \geq 1 - \delta$



Motivations to support the importance of the problem

Understanding the complexity of IRL

Proposing an estimation algorithm with worst case guarantees atop which IRL algorithms can be devised

State of the Art

- **Main Related Works**
  - **Lower and Upper Bounds**
  - **Sample complexity in Bandits**
  - **Sample complexity in Reinforcement Learning**
  - **Sample Complexity in Inverse Reinforcement Learning**
- Limitations
  - Limitations of the works in IRL

Lower and Upper Bounds

Lower Bound minimum number of samples of any algorithm in a certain *difficult* instance

Upper Bound maximum number of samples of the proposed algorithm in any instance

Sample complexity in Bandits

Upper bound $O(\frac{|\mathcal{A}|}{\epsilon^2}\log\frac{1}{\delta})$ in (Even-Dar, Mannor, and Mansour 2002)

Lower bound $\Omega(\frac{|\mathcal{A}|}{\epsilon^2}\log\frac{1}{\delta})$ through the *Likelihood ratio method* (Mannor et al. 2004)

Sample complexity in Reinforcement Learning

Generative model matching bound of $\Theta(\frac{|\mathcal{S}||\mathcal{A}|\bar{H}^3}{\epsilon^2} \log \frac{|\mathcal{S}||\mathcal{A}|}{\delta})$ (Azar, Munos, and Kappen 2012)

Forward model almost matching $O(\frac{|\mathcal{S}|^2|\mathcal{A}|H^2}{\epsilon^2} \log \frac{1}{\delta})$ and $\Omega(\frac{|\mathcal{S}||\mathcal{A}|H^2}{\epsilon^2} \log \frac{1}{\delta + c})$ (Dann and Brunskill 2015)

Sample Complexity in Inverse Reinforcement Learning

Generative model *upper* bound of $\tilde{O}(\frac{|\mathcal{S}||\mathcal{A}|\bar{H}^4}{\epsilon^2})$ samples (Metelli et al. 2021)

Forward model *upper* bound of $\tilde{O}(\frac{|\mathcal{S}||\mathcal{A}|H^5}{\epsilon^2})$ episodes (Lindner, Krause, and Ramponi 2022)

The Lower Bounds?

State of the Art

- Main Related Works
  - Lower and Upper Bounds
  - Sample complexity in Bandits
  - Sample complexity in Reinforcement Learning
  - Sample Complexity in Inverse Reinforcement Learning
- **Limitations**
  - **Limitations of the works in IRL**

Limitations of the works in IRL

No Lower Bound for the generative model

No Lower Bound for the forward model

No distance between feasible sets is considered in the sample complexity

Research Goals

- Nature of the research
- Research goals

Nature of the Research

The research is mostly **theoretical**, a mathematical proof has to be devised

For the upper bound, an algorithm must be proposed and it might be **empirically** evaluated

Research goals

Generative model prove lower and upper bounds

Forward model prove lower and upper bounds

Research Plan

1. study the mathematical tools used in the research topic (March-April);
2. explore the literature (April-June);
3. try to re-use the results for the forward RL problem (June-September);
4. try to prove the bounds in a different way (September-October);
5. devise an algorithm, prove its upper bound and experimentally validate it (October-November);
6. refine the results (November-December);
7. prepare a presentation/paper to present the results (December-January).

Results Obtained so far

- Generative model
- Forward model

Generative model

Lower Bound $\Omega\Big(\frac{|\mathcal{S}||\mathcal{A}|\bar{H}^2}{\epsilon^2}\ln\frac{1}{\delta}\Big)$

Upper Bound $O\Big(\frac{|\mathcal{S}||\mathcal{A}|\bar{H}^2}{\epsilon^2}\ln\frac{1}{\delta}\Big)$ with the proposed algorithm

They match!

Forward model

Lower bound $\Omega\Big(\frac{|\mathcal{S}||\mathcal{A}|H}{\epsilon^2}\ln\frac{1}{\delta}\Big)$

Upper bound $O\Big(\frac{|\mathcal{S}|^2|\mathcal{A}|^2H}{\epsilon^2}\ln\frac{1}{\delta}\Big)$ with the proposed algorithm

Almost matching!

Future Work

$\longrightarrow$ **Refine** the bound for the forward model

$\longrightarrow$ **Understand** the limits of the objective between feasible sets

References I

Azar, Mohammad Gheshlaghi, Munos, Remi, and Kappen, Bert (2012). *On the Sample Complexity of Reinforcement Learning with a Generative Model*. DOI: 10.48550/ARXIV.1206.6461. URL: https://arxiv.org/abs/1206.6461.

Dann, Christoph and Brunskill, Emma (2015). *Sample Complexity of Episodic Fixed-Horizon Reinforcement Learning*. DOI: 10.48550/ARXIV.1510.08906. URL: https://arxiv.org/abs/1510.08906.

Even-Dar, Eyal, Mannor, Shie, and Mansour, Yishay (July 2002). "PAC Bounds for Multi-armed Bandit and Markov Decision Processes". In: pp. 193–209. ISBN: 978-3-540-43836-6. DOI: 10.1007/3-540-45435-7_18.

Haussler, David (1990). "Probably Approximately Correct Learning". In.

Kakade, Sham (Jan. 2003). "On the Sample Complexity of Reinforcement Learning". In.

References II

- Lindner, David, Krause, Andreas, and Ramponi, Giorgia (2022). *Active Exploration for Inverse Reinforcement Learning*. DOI: 10.48550/ARXIV.2207.08645. URL: https://arxiv.org/abs/2207.08645.
- Mannor, Shie et al. (July 2004). "The Sample Complexity of Exploration in the Multi-Armed Bandit Problem". In.
- Metelli, Alberto Maria et al. (2021). "Provably Efficient Learning of Transferable Rewards". In: *Proceedings of the 38th International Conference on Machine Learning*. Ed. by Marina Meila and Tong Zhang. Vol. 139. Proceedings of Machine Learning Research. PMLR, pp. 7665–7676. URL: https://proceedings.mlr.press/v139/metelli21a.html.
- Ng, Andrew Y. and Russell, Stuart J. (Nov. 26, 2002). "Algorithms for Inverse Reinforcement Learning.". In: *ICML*. Ed. by Pat Langley. Morgan Kaufmann, pp. 663–670. ISBN: 1-55860-707-2. URL: http://dblp.uni-trier.de/db/conf/icml/icml2000.html#NgR00.

References III

- Osa, Takayuki et al. (2018). "An Algorithmic Perspective on Imitation Learning". In: *Foundations and Trends in Robotics* 7.1-2, pp. 1–179. DOI: 10.1561/2300000053. URL: https://doi.org/10.1561%2F2300000053.
- Russell, Stuart and Norvig, Peter (2010). *Artificial Intelligence: A Modern Approach*. 3rd ed. Prentice Hall.
- Sutton, Richard S. and Barto, Andrew G. (2018). *Reinforcement Learning: An Introduction*. Second. The MIT Press. URL: http://incompleteideas.net/book/the-book-2nd.html.