INTERNATIONAL JOURNAL OF ARTIFICIAL INTELLIGENCE (ISSN: 2692-515X)
Volume 03, Issue 04, 2023
Published Date: - 09-10-2023

STRATEGIES FOR MITIGATING OVERFITTING AND ASYMPTOTIC BIAS IN BATCH REINFORCEMENT LEARNING WITH PARTIAL OBSERVABILITY

Vincent Pineau
School of Computer Science, McGill University, University Street, Montreal, Canada

Raphael Rabusseau
Montefiore Institute, University of LiegeAllée de la découverte, Belgium

Abstract

In the realm of batch reinforcement learning, where agents are confronted with the challenges of partial observability, the phenomena of overfitting and asymptotic bias can greatly affect the learning process. This paper explores effective strategies for addressing these issues and enhancing the performance of agents operating under partial observability. We delve into novel techniques that provide insights into mitigating overfitting while countering the inherent asymptotic bias, thus paving the way for more robust and reliable batch reinforcement learning algorithms.

Key Words

Batch Reinforcement Learning; Partial Observability; Overfitting; Asymptotic Bias; Mitigation Strategies; Learning Algorithms; Robustness.

INTRODUCTION

Reinforcement learning, a prominent paradigm in the field of machine learning, has demonstrated remarkable success in a wide range of applications, from game playing to robotics and autonomous systems. However, the traditional settings of reinforcement learning, where agents have full access to their environment, do not always hold in practical scenarios. In real-world applications, many systems exhibit partial observability, where agents can only perceive a limited and often noisy subset of the complete state of the environment. This partial observability poses a substantial challenge to the learning process, as agents must make decisions based on incomplete information.

In the context of partial observability, batch reinforcement learning is a valuable approach, which involves learning from a fixed dataset of past experiences. This methodology is particularly relevant in situations where collecting new data for online learning is impractical or costly. However, the success of batch reinforcement learning in partially observable environments is hindered by two critical challenges: overfitting and asymptotic bias.

Overfitting refers to the phenomenon where a learning algorithm becomes too closely tailored to the specific dataset on which it was trained, failing to generalize effectively to new, unseen situations. This issue is exacerbated in batch reinforcement learning with partial observability, as the limited and potentially noisy data can lead to models that are overly sensitive to idiosyncrasies in the training data, hindering their ability to make accurate decisions in real-world scenarios.

Asymptotic bias, on the other hand, arises due to the limited representativeness of the batch data. Since the dataset is finite, learning algorithms may converge towards biased policies that may not approximate the true optimal policy. This bias can have significant implications for the performance and reliability of reinforcement learning agents in partially observable environments.

This paper is dedicated to addressing these pivotal challenges in batch reinforcement learning with partial observability. We present a set of innovative strategies and techniques designed to mitigate overfitting and reduce asymptotic bias. By doing so, we aim to empower reinforcement learning agents to perform more effectively and robustly in scenarios where partial observability is a fundamental characteristic of the environment. These strategies offer a path forward for improving the applicability of batch reinforcement learning in real-world, partially observable domains, ultimately advancing the capabilities of autonomous systems and intelligent agents.

METHOD

In the realm of reinforcement learning, the challenges of overfitting and asymptotic bias loom prominently, particularly when dealing with partially observable environments. Our research explores a multifaceted approach to mitigate these critical issues, paving the way for more robust and reliable batch reinforcement learning with partial observability. By augmenting the dataset, incorporating regularization techniques, and optimizing network architectures, we aim to reduce overfitting while enabling reinforcement learning agents to adapt to a broader range of scenarios. Simultaneously, we employ importance sampling and reweighting techniques to combat asymptotic bias, ensuring that learned policies approach a more accurate representation of the true optimal policy. Rigorous evaluation metrics and benchmarks are introduced to assess agent performance accurately. Through this comprehensive strategy, we aspire to empower reinforcement learning agents to excel in real-world scenarios characterized by partial observability, further advancing the capabilities of intelligent systems in complex and dynamic environments.

The process of developing and implementing the strategies for mitigating overfitting and asymptotic bias in batch reinforcement learning with partial observability involves a series of well-defined steps. These steps are carefully orchestrated to ensure that the resulting framework enhances the robustness and performance of reinforcement learning agents in complex, partially observable environments.

Problem Formulation and Data Collection:

The first step in our process involves formulating the reinforcement learning problem in the context of partial observability. We define the state space, action space, and rewards, while acknowledging that the agent has limited access to the environment. A batch of historical data is collected, representing past experiences of the agent within the partially observable environment.

Augmented Dataset Generation:

To address overfitting, we augment the original batch dataset. This augmentation process involves generating additional experiences, both by adding noise to existing data and by simulating new data points. These artificial experiences introduce diversity into the dataset, allowing the agent to learn from a broader range of scenarios.

Regularization and Network Architecture Selection:

We introduce regularization techniques to the learning process to control overfitting. This includes applying L1 and L2 regularization, dropout, and data augmentation. Additionally, we carefully select network architectures for the policy and value functions. Recurrent neural networks (RNNs) and attention mechanisms are explored to capture temporal dependencies in observations and improve information fusion under partial observability.

Importance Sampling and Reweighting:

To mitigate asymptotic bias, we incorporate importance sampling and reweighting techniques into the learning algorithm. These methods involve assigning appropriate importance weights to experiences in the dataset, correcting for the inherent bias in finite and potentially non-representative data.

Evaluation and Benchmarking:

A crucial aspect of our process is the development of specialized evaluation metrics and benchmarks. These metrics go beyond traditional performance measures, accounting for the challenges of partial observability. We rigorously evaluate the performance of reinforcement learning agents against these benchmarks to assess their capabilities.

Hyperparameter Tuning:

The process also includes an extensive hyperparameter search, where we fine-tune the configuration of the learning algorithm. This tuning phase aims to strike the right balance between overfitting and asymptotic bias mitigation, optimizing the agent's performance in the given application.

Experimentation and Validation:

The strategies and techniques developed are put to the test through extensive experimentation. Reinforcement learning agents are trained and evaluated in partially observable environments, and their performance is compared to baseline methods. The experiments validate the effectiveness of our strategies in reducing overfitting and asymptotic bias.

Iterative Refinement:

The process often involves iterative refinement, where the strategies and techniques are adjusted and improved based on the results of experimentation. Feedback from the evaluation phase informs further enhancements to the framework.

Through this structured process, we aim to provide a comprehensive solution for addressing the challenges of overfitting and asymptotic bias in batch reinforcement learning with partial observability. The resulting framework is designed to empower reinforcement learning agents to navigate complex and partially observable environments more effectively, advancing the state of the art in autonomous systems and intelligent agents.

RESULTS

The experimental results of our strategies for mitigating overfitting and asymptotic bias in batch reinforcement learning with partial observability demonstrate significant improvements in agent performance and reliability. When compared to baseline methods, the reinforcement learning agents trained using our approach exhibit reduced overfitting and a more accurate approximation of the true optimal policy. The augmented dataset, regularization techniques, and specialized network architectures collectively contribute to a substantial enhancement in the agent's capacity to handle partial observability.

The introduction of importance sampling and reweighting techniques effectively reduces asymptotic bias in the learned policies. Our agents converge toward policies that better approximate the true optimal policy, improving their decision-making capabilities in partially observable environments.

DISCUSSION

The success of our strategies can be attributed to the synergy between various components in our approach. Augmenting the dataset allows the agent to learn from a more diverse range of experiences, mitigating overfitting by reducing sensitivity to idiosyncrasies in the training data. Regularization techniques help in constraining the model's complexity, preventing it from fitting noise in the data and encouraging generalization.

The choice of policy and value function architectures, specifically incorporating recurrent neural networks and attention mechanisms, enhances the agent's ability to capture temporal dependencies and fuse information in a partially observable setting. This proves crucial in scenarios where the agent's actions are influenced by a history of observations.

The application of importance sampling and reweighting techniques rectifies the bias that arises due to the finite and potentially non-representative nature of the batch dataset. By assigning appropriate importance weights to experiences, we ensure that the agent's learned policies approach more accurate approximations of the true optimal policy.

CONCLUSION

In conclusion, our research has contributed valuable strategies for addressing the challenges of overfitting and asymptotic bias in batch reinforcement learning with partial observability. The combination of dataset augmentation, regularization techniques, network architecture selection, importance sampling, and reweighting, as well as the use of specialized evaluation metrics, has proven to be effective in enhancing the performance and reliability of reinforcement learning agents in complex, partially observable environments.

These strategies offer a promising path forward for the deployment of reinforcement learning agents in real-world applications, where the full observability of the environment cannot be guaranteed. By reducing overfitting and asymptotic bias, our approach empowers agents to make more accurate and robust decisions in dynamically changing and uncertain environments. This, in turn, advances the capabilities of autonomous systems and intelligent agents, making them more adaptable and dependable in a wide range of practical scenarios. Our work underscores the importance of developing strategies that can successfully navigate the perils of overfitting and asymptotic bias, ultimately contributing to the broader field of reinforcement learning and its practical applications.

REFERENCES

1. Abel, D., Hershkowitz, D., & Littman, M. (2016). Near optimal behavior via approximatestate abstraction. InProceedings of The 33rd International Conference on MachineLearning, pp. 2915–2923.

2. Aberdeen, D. (2003). A (revised) survey of approximate methods for solving partiallyobservable Markov decision processes.National ICT Australia, Canberra, Australia, 1.

3. Aberdeen, D., Buffet, O., & Thomas, O. (2007). Policy-gradients for PSRs and POMDPs.InArtificial Intelligence and Statistics, pp. 3–10.

4. Arun-Kumar, S. (2006). On bisimilarities induced by relations on actions. InSoftwareEngineering and Formal Methods, 2006. SEFM 2006. Fourth IEEE International Conference on, pp. 41–49. IEEE.

5. Cassandra, A. R., Kaelbling, L. P., & Littman, M. L. (1994). Acting optimally in partially observable stochastic domains. InProceedings of the Twelfth AAAI National Conferenceon Artificial Intelligence, Vol. 94, pp. 1023–1028.

6. Castro, P. S., Panangaden, P., & Precup, D. (2009). Equivalence relations in fully andpartially observable markov decision processes.. InTwenty-First International JointConference on Artificial Intelligence, Vol. 9, pp. 1653–1658.

7. Chen, T., Goodfellow, I., & Shlens, J. (2015). Net2net: Accelerating learning via knowledgetransfer.arXiv preprint arXiv:1511.05641.

8. Farahmand, A.-m. (2011).Regularization in reinforcement learning. Ph.D. thesis, Universityof Alberta.

9. Ferns, N., Panangaden, P., & Precup, D. (2004). Metrics for finite Markov decision processes.InProceedings of the 20th conference on Uncertainty in artificial intelligence, pp.162–169. AUAI Press.

10. François-Lavet, V., Fonteneau, R., & Ernst, D. (2015). How to discount deep reinforcementlearning: Towards new dynamic strategies.arXiv preprint arXiv:1512.02011.

11. François-Lavet, V., Taralla, D., Ernst, D., & Fonteneau, R. (2016). Deep reinforcementlearning solutions for energy microgrids management. In European Workshop onReinforcement Learning.

12. Ghavamzadeh, M., Mannor, S., Pineau, J., & Tamar, A. (2015). Bayesian reinforcementlearning: a survey.Foundations and Trends®in Machine Learning,8(5-6), 359–483.