Department of Electrical Engineering and
Information Technology
Automatic Control Group
https://en.ei.uni-paderborn.de/de/rat/

BACHELORARBEIT

Student-Teacher vs. Teacher-Student Learning: Wer sollte schneller lernen für besseren gemeinsamen Erfolg?

Hintergrund

Im bestärkenden Lernen gibt es eine verbreitete Methode, die man vereinfacht Student-Teacher Learning nennen könnte. Dabei soll ein Actor/Student lernen, in einer Umgebung möglichst gute Entscheidungen zu treffen. Die Entscheidungen haben dabei nicht nur direkte Konsequenzen, sondern auch langfristige und beeinflussen ebenfalls die Situationen, die der Actor in der Zukunft möglicherweise erlebt. Der Actor interagiert also mit einem dynamischen System und das Ziel ist es, eine möglichst gute Entscheidungsstrategie für den Actor zu finden. Der Actor tut dies allerdings nicht alleine, sondern kann auf die Hilfe vom Critic/Teacher setzen. Der Critic lernt aus dem Verhalten des Actor im System laufend die Entscheidungen des Actors zu evaluieren. Basierend auf der Evaluation des Critics, ändert der Actor dann sein Verhalten. Die prinzipielle Struktur dieses Lernalgorithmus ist in Abbildung 1 dargestellt.

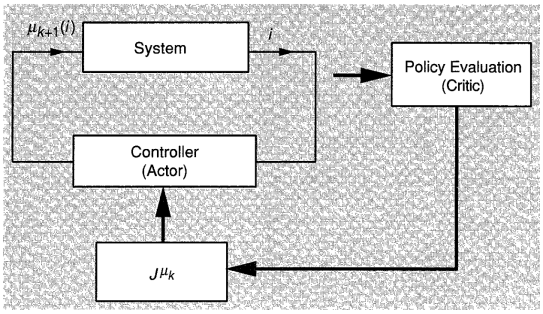

Abbildung 1: Struktur vom Actor-Critic Learning für Strategie/Policy μ_k mit Evaluation J^{μ_k} .

Ziel der Arbeit

Bei der Implementierung des Actor-Critic Lernalgorithmus stellt sich nun eine fundamentale Frage:

Wer soll schneller Lernen? Der Actor(Student) oder der Critic(Teacher)?

Das Problem ist Folgendes: Wenn der Actor seine Strategie stark ändert, braucht der Critic zunächst möglicherweise lange, um die langfristige Auswirkung der Strategieänderung des Actors gut zu evaluieren. Aufgrund dieser Intuition wurde in der bisherigen Literatur der Actor immer so implementiert, dass er "langsamer" lernt als der Critic [1]. Der Critic sollte Änderungen der Actorstrategie schnell evaluieren können.

Kürzlich hat sich jedoch gezeigt, dass es aus theoretischer Sicht keinen wesentlichen Grund gibt, den Critic schneller als den Actor lernen zu lassen [2]. *Bisher gibt es jedoch keine Antwort darauf, wann Actor-Critic Learning (schnellerer Critic) oder Critic-Actor Learning (schnellerer Actor) zu empfehlen ist.* Ziel dieser Arbeit ist es, in einer einfachen Minigrid Lernumgebung zu testen, wann Actor-Critic Learning oder Critic-Actor Learning zu besserem Lernfortschritt führt.

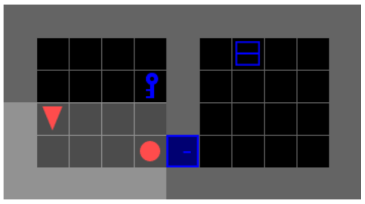

Abbildung 2: Minigrid Umgebung in der ein Actor Lernen muss mit einem Schlüssel eine zunächst blockierte Tür zu öffnen, um eine Box zu erreichen [3].

Meilensteine

- Einarbeitung in Actor-Critic Learning und die Minigrid Simulationsumgebung.
- Implementierung von Actor-Critic Learning und testen mit verschiedenen Lernraten.
- Vergleich von Actor-Critic and Critic-Actor Learning in verschiedenen Minigrid Umgebungen.
- Evaluation und Dokumentation.

Voraussetzungen

- Programmieren in Python.
- Grundlagen der Systemtheorie.

Literatur

[1] V. Konda and J. Tsitsiklis, "Actor-critic algorithms," *Advances in neural information processing systems*, vol. 12, 1999.

[2] S. Bhatnagar, V. S. Borkar, and S. Guin, "Actor-critic or critic-actor? a tale of two time scales," *IEEE Control Systems Letters*, 2023.

[3] M. C.-B. et al., "Minigrid & miniworld: Modular & customizable reinforcement learning environments for goal-oriented tasks," *CoRR*, vol. abs/2306.13831, 2023.

Contact: Adrian Redder, aredder@mail.upb.de, Room: P1.7.09.2