



Universiteit
Leiden
The Netherlands

Enhancing autonomy and efficiency in goal-conditioned reinforcement learning

Yang, Z.

Citation

Yang, Z. (2025, February 26). *Enhancing autonomy and efficiency in goal-conditioned reinforcement learning*. SIKS Dissertation Series. Retrieved from <https://hdl.handle.net/1887/4196074>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/4196074>

Note: To cite this publication please use the final published version (if applicable).

Samenvatting

Reinforcement learning is een methode om te leren door vallen en opstaan, en dat werkt voor computers net als bij mensen. Idealiter willen we een generiek algoritme hebben dat een agent zodanig traint dat deze verschillende taken kan uitvoeren en verschillende doelen kan bereiken. Goal-directed reinforcement learning is een manier die een stap in deze richting zet. Het Goal-directed reinforcement learning raamwerk omvat vier stappen: 1) het definiëren van de doelruimte; 2) het selecteren van een interessant doel voor de agent; 3) een algoritme om de agent te leren om dat doel te bereiken; 4) een agent die post-exploration uitvoert. In het proefschrift poneren we vier onderzoeksvragen. We richten ons op drie delen van het goal-directed reinforcement learning raamwerk. Ten eerste hebben we in Chapter 3 een meer autonoom goal-directed reinforcement learning setting bestudeerd die werkt zonder reset; dit levert complicaties op bij exploratie. Door zogenaamde World-Models te gebruiken en verschillende doelen te selecteren in de fase van gegevensverzameling, kan onze voorgestelde agent autonoom werken in reset-vrije taken en beter presteren dan state-of-the-art baselines. Zodra een doel is geselecteerd, wordt normaalgesproken een goal-directed policy getraind om het doel te bereiken. Eerder onderzoek naar een niet-parametrische methode (genaamd model-free episodic control (Blundell, Uria, Pritzel, Li, Ruderman, Leibo, Rae, et al., 2016)), toont aan dat episodic control snel kan inhaken op eerder ontdekte oplossingen en sneller kan leren dan deep RL-methoden (waarvan bekend is dat ze traag zijn). In plaats van het trainen van een goal-directed policy, kan episodic control ook worden gebruikt om de agent te trainen om het geselecteerde doel te bereiken. Model-free episodic control is echter ontworpen voor taken met een discrete actieruimte. In Chapter 4 breiden we episodic control uit naar continue episodische methoden die nu taken met een continue actieruimte kunnen aanpakken. We laten zien dat dit goed presteert op een aantal continue problemen uit de robotica. Ook noemen we een aantal beperkingen van episodic control, in Chapter 5, namelijk dat ze niet optimaal zullen

Samenvatting

zijn in stochastische problemen. Vervolgens stelden we voor om episodisch control te combineren met diep reinforcement learning-methoden om van het beste van twee werelden te bereiken. Onze experimenten laten zien dat door deze twee methoden te combineren, de gecombineerde agent zowel het snelle leren bereikt van episodisch control, als de optimaliteit van reinforcement learning. In Chapter 6 laten we zien dat extra exploratie na het bereiken van de doelen op de search-boundary (post-exploratie genoemd) de efficiëntie van GCRL-agenten verbetert. Zonder post-exploratie reset de agent zich na het bereiken van deze doelen, waardoor de agent zijn kennisgrens niet kan uitbreiden. Met post-exploratie stappen agenten effectief in nieuwe, onbekende, gebieden en verwerven ze meer diverse gegevens. Dit leidt tot betere prestaties in vergelijking met agenten die geen post-exploratie uitvoeren. De methoden die in het proefschrift worden voorgesteld, verbeterden onderdelen van het goal-directed reinforcement learning-framework, waaronder doelselectie, policy-learning en exploratie, wat heeft geleid tot betere prestaties en de autonomie van het gehele goal-directed reinforcement learning-framework verhoogde. In de toekomst hopen we dat goal-directed reinforcement learning gebruikt zal worden voor het trainen van een generieke agent en uiteindelijk een embodied AI agent (robot).