Searching by learning: Exploring artificial general intelligence on small board games by deep reinforcement learning

Wang, H.

Nederlandse Samenvatting

Twee belangrijke componenten in het vakgebied Deep Reinforcement Learning zijn technieken om te zoeken en te leren. Ze kunnen los van elkaar en in combinatie worden gebruikt om uiteenlopende problemen in de kunstmatige intelligentie op te lossen, en hebben hierin indrukwekkende resultaten bereikt, zowel in robotica als in het spelen van spellen. Deze resultaten hebben onderzoek naar kunstmatige *brede* intelligentie met behulp van zoek- en leertechnieken gestimuleerd (artificial general intelligence: AGI). Twee benaderingen—General Game Playing (GGP) en AlphaZero—worden veel gebruikt om verschillende aspecten van AGI te onderzoeken. Deze twee benaderingen gebruiken combinaties van zoek- en leertechnieken.

Het doel van dit proefschrift is om te kijken hoe ver we kunnen komen met deze technieken. We bestuderen het klassieke Q-learning algoritme (dat nog een tabel als basis-datastructuur gebruikt) in een GGP systeem. We tonen hiermee aan dat klassiek Q-learning werkt in GGP, al convergeert het langzaam, en is er erg veel rekenkracht nodig om ingewikkelder spellen te leren spelen. Voor grotere spellen zouden diepe neurale netwerken wel eens beter kunnen werken, hetgeen ons volgende experiment is. Bestaand onderzoek geeft aan dat de combinatie van zoeken en leren beter zal presteren. In deze werken maken zoektechnieken gebruik van neurale netwerken die getraind worden met voorbeelden die het zoekalgoritme dan weer aanlevert—de zogeheten *self-play* aanpak. In dit proefschrift gebruiken we een op AlphaZero gebaseerde benadering om AGI in kleine spellen te onderzoeken. We onderzoeken met verschillende waarden voor hyperparameters de rol, het effect, en de bijdrage van zoek- en leertechnieken. Teneinde het relatieve belang van zoeken en leren te bekijken, hebben we een experiment gedaan met als uitkomst dat het aantal zoek-iteraties van groter belang is dan training-iteraties, simulaties, of spel-episoden, dan tot nog toe gedacht.

Een ander experiment laat zien dat zoektechnieken gebruikt kunnen worden als expert om de start van het leerproces te verbeteren. Dit idee noemen we de *warme*

start in dit proefschrift. In de AlphaZero benadering kan een combinatie van Rollout en RAVE technieken de eerste iteraties van self-play-training verbeteren, en helemaal met een adaptieve iteratie-lengte. De warme-start techniek is een veelbelovende trainingsmethode om toe te voegen aan de self-play aanpak.

Na het succes van AlphaZero self-play methoden in spellen voor twee personen, onderzoeken we of dit succes ook gebruikt kan worden in spellen voor één persoon, ofwel puzzels. Hiertoe implementeren we het spel Morpion Solitaire met de Ranked Reward methode. Morpion Solitaire is een heel ingewikkelde combinatorische puzzel. Onze eerste AlphaZero-achtige aanpak bleek al in staat om het beste menselijke record te benaderen. Dit geeft aan dat dit een veelbelovende AGI benadering is voor eenpersoons spellen.

In dit proefschrift worden zoek- en leertechnieken bestudeerd, zowel apart als in combinatie, in GGP en in AlphaZero-achtige self-play systemen. Ons doel is om stappen te zetten om kunstmatige intelligentie breder toepasbaar te maken, naar systemen die intelligent gedrag vertonen in meer dan een domein. Onze resultaten zijn veelbelovend, en laten zien in welke richting zoektechnieken kunnen worden toegepast om leertechnieken te verbeteren.