Reinforcement Learning en Computerspellen

Hoe beïnvloeden de specifieke kenmerken van computerspellen de effectiviteit van specifieke reinforcement learning-algoritmes?

Christelijk Lyceum Zeist

Matthijs Gorter
Thom Brinkhorst
Pepijn van Iperen

Profielwerkstuk
onder begeleiding van
S. Rook
Christelijk Lyceum Zeist
Natuur en Techniek
Februari 2025

Voorwoord

Voorwoord inhoud hier. Bedank mensen die geholpen hebben, beschrijf het doel van het profielwerkstuk en eventuele persoonlijke motieven of ervaringen.

Matthijs Gorter, Thom Brinkhorst, Pepijn van Iperen
Christelijk Lyceum Zeist
Februari 2025

Variabelen en Notatie

Variabele	**Definitie**
t	Tijdstap
T	Laatste tijdstap van een episode (horizon)
x	Toestand (state)
x_t	Toestand op tijdstip t
x'	Toestand een tijdstap na x
$\mathcal{X}$	Set van alle toestanden
a	Actie
$\mathcal{A}$	Alle mogelijke acties
a_t	Actie op tijdstip t
r	Beloning (reward)
$\mathcal{R}$	Set van mogelijke beloningen
r_t	Beloning op tijdstip t
$r(x, a)$	Beloningsfunctie
μ	Deterministisch beleid
π	Stochastisch beleid
π^*	Optimale stochastisch beleid
γ	Kortingsfactor tussen 0 en 1
$p(x'\|x, a)$	Overgangswaarschijnlijkheidsfunctie
$\mathcal{P}$	Overgangswaarschijnlijkheidsmatrix
$V(x)$	Waardefunctie
$Q(x, a)$	Q-functie
$Q^*(x, a)$	Q-functie met het optimale beleid
$\mathbb{E}[X]$	Verwachtingswaarde van variabele X
$\mathbb{E}[a\|b]$	Geconditioneerde verwachtingswaarde
$\mathbb{E}_\pi[X]$	Verwachtingswaarde als beleid π wordt gevolgd

Tabel 1: Variabelen en Notatie

Inhoudsopgave

1

Inleiding

Introductie Onderwerp
Onderwerpkeuze verantwoorden
Onderzoeksvraag/Hoofdvraag met eventuele hypothese
Deelvragen
Wat zijn de specifieke kenmerken van verschillende soorten computerspellen?
Welke reinforcement learning-algoritmes zijn beschikbaar en wat zijn hun kenmerken?
Hoe beïnvloeden de spelkenmerken de prestatie van reinforcement learning-algoritmes?
klein stukje theorie als inleiding op het theorie-onderdeel
werkplan in grote lijnen, opbouw van verslag

2

Theoretisch Kader

2.1 Definitie

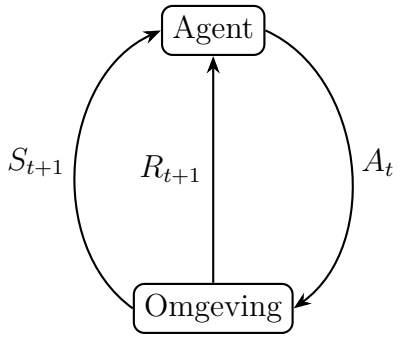
Reinforcement Learning (RL) is een tak binnen kunstmatige intelligentie waarin een agent leert door interactie met zijn omgeving. Een agent is een entiteit die leert en acties onderneemt. Bij een zelfrijdende auto is het besturingssysteem de agent, en bij een schaakspel is de schaker de agent. De omgeving is alles waarmee de agent interageert en die reageert op de acties van de agent. Bij een zelfrijdende auto is dit de weg waar de auto op rijdt en de voertuigen om de auto heen. Bij een schaakspel is dit het schaakbord.

De agent leert door interactie met zijn omgeving. De agent ontvangt beloningen of straffen (negatieve beloningen) als gevolg van zijn acties. Het doel van de agent is om een strategie te ontwikkelen die de cumulatieve beloning maximaliseert over tijd. Bij Super Mario Bros (1985) is Mario de agent en de agent krijgt een beloning als Mario richting de eindslag beweegt (meestal naar rechts) en als Mario een coin of een power up oppakt en Mario krijgt een straf als Mario sterft en hij krijgt elke seconde straf zodat hij zo snel mogelijk het level wilt voltooien.

Figuur 2.1 illustreert het basismodel van interactie binnen Reinforcement Learning.

Een actie naar de beslissing die een agent neemt bij elke stap in een besluitvormingsproces. Acties worden aangeduid met a en worden gekozen uit een reeks mogelijke acties $\mathcal{A}$. Elke door de agent genomen actie beïnvloedt de interactie met de omgeving, wat leidt tot een verandering in de toestand en een daaruit voortvloeiende beloning.

Een toestand x vertegenwoordigt de huidige situatie of staat van de omgeving waarin



Figuur 2.1: Reinforcement learning interactiemodel tussen agent en omgeving via acties, toestanden en beloningen.

de agent opereert. Dit wordt aangeduid met x en maakt deel uit van de toestandsruimte $\mathcal{X}$. Bij de aanvangsstap $t = 0$, begint de agent in een initiële toestand x_0 die willekeurig wordt bepaald door een verdeling p. Naarmate het proces vordert, bevindt de agent zich in nieuwe toestanden gebaseerd op zijn acties.

Een beloning r is een feedbackwaarde die wordt ontvangen nadat de agent een actie heeft uitgevoerd in een bepaalde toestand. Deze beloning wordt bepaald door de beloningsfunctie $r(x, a)$. De beloningsmatrix R bevat de onmiddellijke beloningen voor elke combinatie van toestand en actie.

Een overgang beschrijft de verandering van de huidige toestand naar de volgende toestand als gevolg van een actie die door de agent wordt genomen. De waarschijnlijkheid van overgang wordt bepaald door de overgangswaarschijnlijkheidsfunctie $p(x'|x, a)$, die afhangt van de huidige toestand x, de genomen actie a en leidt tot een nieuwe toestand x'. De overgangswaarschijnlijkheidsmatrix P bevat de waarschijnlijkheden van het overgaan van de ene toestand naar de volgende toestand, gegeven een bepaalde actie.

2.2 Belangrijke Algoritmes

Algorithm 1 Example Algorithm

1: **Input:** An array A of n integers
2: **Output:** The sum of the integers in A
3: $sum \leftarrow 0$
4: **for** $i \leftarrow 1$ to n **do**
5: $\quad sum \leftarrow sum + A[i]$
6: **end for**
7: **return** sum

2.3 Artificiële Leermethodes

2.4 Computerspellen

3

Onderzoeksmethoden

4

Analyse en Resultaten

5

Conclusie

6

Discussie

7

Referentielijst

8

Bijlagen