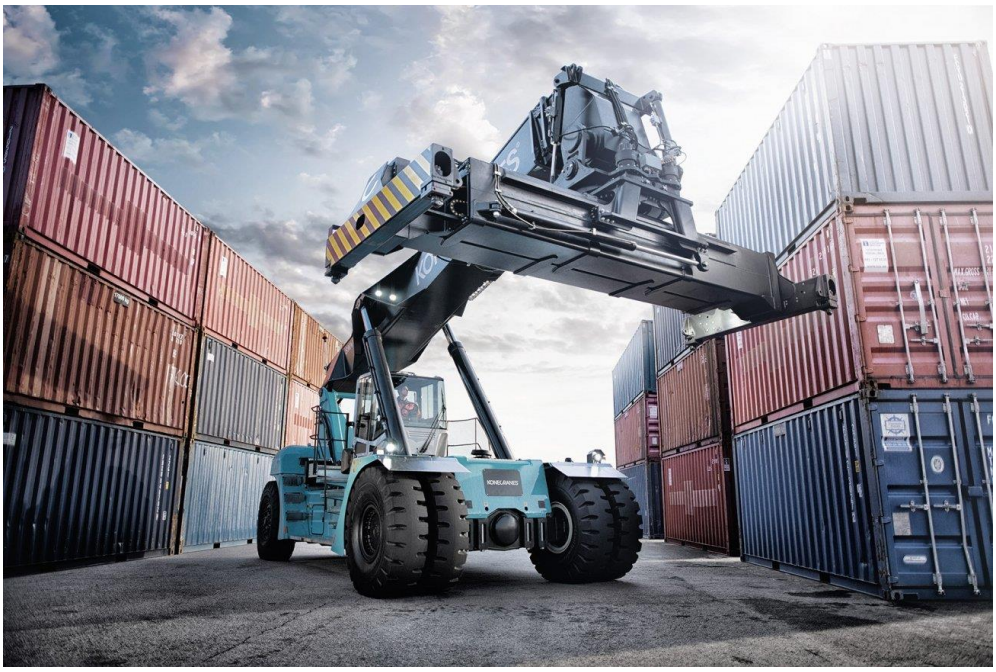
REINFORCEMENT LEARNING MODEL



In opdracht van De Haagse Hogeschool
24 januari 2023

Studenten Projectgroep 1:

Bonno van Nieuwenhout	19122381
Hidde Franke	19086504
Joeri Meijers	20123558
Manon Rongen	19075235
Michael Broer	20105533
Mohamed Amajoud	20198752

Begeleid door:

Jeroen Vuurens
Tony Andrioli
Karin de Smidt - Destombes
Edwin van Noort

Samenvatting

Dit onderzoek richt zich op het zo optimaal mogelijk indelen van yards in containerterminals door middel van geautomatiseerde methodieken. Het bedrijf Cofano, dat zich bezighoudt met het optimaliseren van terminalprocessen, streeft ernaar om de tijd dat schepen aan de kade liggen te minimaliseren en daarmee de kosten zo laag mogelijk te houden. Eén van deze terminalprocessen wordt gedaan door middel van reach stackers, die containers verplaatsen vanaf de yard naar schepen voor verder vervoer. Dit proces wordt geoptimaliseerd door het zo efficiënt mogelijk indelen van de yard, waardoor de stackers de containers in een minimaal aantal stappen verplaatsen. Er is een schaalbaar Reinforcement Learning (RL) model ontwikkeld om dit complexe probleem op te lossen. Dit model is uit te breiden naar grotere yards, opstapeling van containers of situaties met meer containers en schepen. Gefocust op deze factoren heeft het model een eindoplossing gecreëerd waarbij de stacker bij alle gewenste containers kan komen, zonder extra stappen. Na evaluatie blijkt dit model schaalbaar te zijn door een optimale oplossing te genereren voor yards met verschillende groottes.

Inhoudsopgave

1. Introductie

Introductie

In dit hoofdstuk zal een korte introductie gegeven worden over het containerbedrijf Cofano en het probleem wat behandeld wordt in dit onderzoek. Vervolgens wordt het onderzoeksdoel omschreven en ten slotte worden er andere werken, die gerelateerd zijn aan dit onderzoek, besproken in het literatuuronderzoek. Na de introductie kan er gelezen worden over de opzet van dit onderzoek, hoe de literatuur is gekoppeld aan dit probleem, de dataverzameling en de methodologie. Daarna volgt het hoofdstuk met de onderzoeksresultaten, waarin de belangrijkste resultaten getoond zullen worden. Ten slotte kunt u de conclusie en discussie lezen, waarin de resultaten teruggekoppeld worden aan de doelstelling van dit onderzoek en mogelijke verbeteringen en/of punten waar tegenaan zijn gelopen gedurende dit onderzoek worden benoemd.

Probleemstelling

Cofano is een bedrijf dat zich bezighoudt met het optimaliseren van terminalprocessen. Deze processen houden het volgende in: er komen binnenvaartschepen aan, die containers af moeten lossen op de kade van een terminal. De containers staan daar vervolgens totdat een zeevaartschip de containers komt halen. In de volgorde waarop schepen de terminal in- en uitvaren zit een onzekerheid, met deze onzekerheid moet rekening gehouden worden. Cofano wil de tijd dat schepen aan de kade liggen minimaliseren om de kosten zo laag mogelijk te houden. Dit betekent dat de in- en aflaadprocessen geoptimaliseerd moeten worden.

Er wordt in de terminals gewerkt met reach stackers. Deze machines zullen verder in dit paper worden aangeduid als “stackers”. Hoe minder stappen een stacker nodig heeft om een container te bereiken, hoe sneller het proces verloopt. Deze stackers kunnen alleen containers pakken vanaf de lange zijde en vanaf verschillende hoogtes. Als eerst moeten bovenste containers verplaatst worden, als een onderste container gepakt moet worden. Als een container tussen andere containers ligt, kan de stacker hier niet bij. Het is mogelijk maximaal vijf containers op elkaar te stapelen. Een visualisatie met een voorbeeld van een onbereikbare container kan gevonden worden in de bijlage, (zie Bijlage, Figuur 7). Het doel van Cofano is om deze verschillende processen te optimaliseren om de kosten zo laag mogelijk te houden.

Het doel van dit onderzoek

In dit onderzoek is een methodologie ontwikkeld voor het automatiseren van het vinden van een optimale indeling van de yard in een terminal, zodat een stacker containers ophaalt voor verder vervoer in een minimaal aantal stappen. Dit is een complex probleem. Er zijn diverse soorten containers, en die zijn van diverse grootte en gewicht. Daarnaast zijn er vele containerterminals in de wereld, en die verschillen qua grootte, aantal containers en soorten containers. Dit maakt de zoektocht naar een algemene oplossing lastig.

Omdat het een dynamisch probleem is, is het van belang dat het model schaalbaar is. Dat wil zeggen dat het makkelijk uit te breiden is naar bijvoorbeeld een grotere yard of een situatie met meer containers en zeevaartschepen. Wel zijn er een aantal aannames gedaan waar het model mee werkt. Het model houdt geen rekening met verschillende soorten containers en gaat uit van één yard waar containers geplaatst kunnen worden, zonder rekening te houden met de verdere indeling van een terminal.

Literatuuronderzoek

Het Container Stacking Problem (CSP) is een bekend probleem, waar al meerdere onderzoeken naar zijn gedaan. In deze onderzoeken wordt er gekeken naar het optimaliseren van bepaalde terminalprocessen. Het probleem is op te delen in losse problemen, zoals bijvoorbeeld het uitladen en plaatsen van containers, het inzetten van stackers en het vervoeren van containers van de terminal naar de zeevaartschepen. Er zijn verschillende heuristieken gebruikt om zulke problemen op te lossen.

Kefi et al. (2007) vergelijkt een Uninformed en Informed Search Algorithm voor het toewijzen van een slot aan een container. Het model minimaliseert het aantal bewegingen en verplaatsingen die nodig zijn om een container te pakken. Het Informed Search Algorithm geeft de meest optimale oplossing voor het probleem. Salido et al. (2009) heeft Artificial Intelligence (AI) toegepast op het CSP. Net als in het onderzoek wat in dit paper beschreven wordt, richt dit model zich op een optimale plaatsing van containers voor het ophaalproces. Het doel is hetzelfde als in dit onderzoek, namelijk het minimaliseren van het aantal verplaatsingen dat nodig is om een container te pakken.

In een ander onderzoek (Ries et al., 2014) wordt een Fuzzy Logic Model gebruikt om binnenkomende containers toe te wijzen aan een plek in de terminal. In dit onderzoek ligt de focus op het bouwen van een flexibel model, dat rekening houdt met onzekerheden van aankomsttijden van containers bij de terminals en wat bruikbaar is voor terminals met verschillende indelingen en infrastructuren. Het is voor dit onderzoek belangrijk te werken aan een flexibel model voor het vinden van een optimale indeling, omdat in dit onderzoek de onzekerheid van aankomst van containers een rol speelt.

Daarnaast is op dit gebied ook al gewerkt met Reinforcement Learning om een optimale oplossingen te vinden voor terminalprocessen. In het onderzoek van Jiang et al. (2021) ligt de focus op het optimaal stapelen van containers op basis van prioriteit van een container. Het doel is om het aantal relocaties dat moet plaatsvinden voordat alle containers op een bepaalde volgorde gepakt kunnen worden, te minimaliseren. Dit probleem wordt aangepakt met RL en de resultaten zijn net zo goed als de meest optimale oplossingen die met andere methodes gevonden zijn.

In het onderzoek van Hu et al. (2021) wordt met twee verschillende methodes, Integer Programming (IP) en RL, het vervoeren van containers naar het zeevaartschip geoptimaliseerd. Het doel is het vinden van een optimale operatievolgorde en optimale routes voor de stackers, die rijden tussen de containers en het zeevaartschip. Het RL-model scoort daarbij het beste. De resultaten van Jiang et al. (2021) en Hu et al. (2021) laten zien dat RL een geschikte methode is om soortgelijke problemen aan te pakken.

Hu et al. (2023) bekijkt het optimaliseren van het planningsproces in containerterminals. Hier wordt een systeem environment gebouwd en door middel van RL gezocht naar een optimale planning van operaties in de terminal. Hier is ondervonden dat RL efficiënt werkt en flexibeler is dan andere heuristieken. Dit is erg belangrijk voor het CSP, aangezien daar veel onzekerheden bij komen kijken en het een dynamisch probleem is.

In het onderzoek van Krishna & Sudhir (2020) worden meerdere experimenten gedaan waarbij RL-modellen worden vergeleken. Dit onderzoek kan toegepast worden voor het zo efficiënt mogelijk indelen van de containers op de kade. Uit onderzoek naar de modellen bleek dat het A2C-model voor een deel gebaseerd is op het PPO-model. Echter zijn de resultaten van het PPO-model voor de experimenten, die besproken zijn in dit onderzoek, beter.

In een video (Renotte, 2021) over RL wordt aandacht besteed aan het zelf bouwen van een environment en RL-model. Stap voor stap wordt een douche environment opgebouwd en een model die de temperatuur gedurende een uur optimaliseert. Met behulp van deze video kan er een vergelijkbaar RL-model gebouwd worden die, bijvoorbeeld voor dit onderzoek, containers op een zo optimaal mogelijke manier plaatst.

2. Onderzoeksopzet

Toepassing van het theoretisch kader

Om een efficiëntere inrichting van containerterminals te bereiken, is er eerst onderzoek gedaan naar bestaande methoden. Op basis van dit onderzoek zal er gericht gekeken worden naar verbeteringen van deze methoden. Uit het literatuuronderzoek blijkt RL een geschikte techniek is om dit probleem aan te pakken. Daarom richt dit onderzoeksopzet zich op het gebruik van RL. Deze benadering is gekozen omdat RL zich leent voor problemen waarbij een agent leert van zijn acties en de gevolgen daarvan in een dynamische omgeving. Bovendien is het een stabiele en robuuste methode voor het oplossen van problemen met veel discrete acties.

Uit het literatuuronderzoek blijkt dat er een aantal modellen geschikt zijn om toe te passen voor het zo efficiënt mogelijk indelen van de kade. Uit onderzoek naar de modellen bleek dat het A2C-model voor een deel gebaseerd is op het PPO-model. Het verschil is dat het A2C-model agressiever zoekt naar een verbetering. Dit is terug te zien in de value loss van het A2C-model vergeleken met het PPO-model, de value loss is een grafiek die laat zien hoe het model leert over de tijd. Bij het A2C-model is dit een heel erg fluctuerende lijn, bij het PPO-model loopt deze lijn een stuk vlakker. Uiteindelijk zijn beide modellen uitgeprobeerd en is er voor dit onderzoek gekozen voor het PPO-model omdat hier de beste resultaten uit voort kwamen.

Dataverzameling

Om het RL-model te trainen en evalueren, is gebruik gemaakt van gesimuleerde data in plaats van de beschikbare data van Cofano, aangezien die data niet bruikbaar was voor dit onderzoek. De gesimuleerde data die gebruikt is voor het trainen en evalueren van dit model zijn gebaseerd op een lijst van containers die zijn toegewezen aan specifieke zeevaartschepen, gedefinieerd door het nummer van de container. Deze lijst is opgebouwd met een gelijke verdeling van containers voor elk zeevaartschip. De volgorde van binnenkomst van containers wordt willekeurig gesorteerd om variatie in de data te simuleren.

De containers worden één voor één vanuit de lijst in een yard geplaatst en vervolgens uit de lijst verwijderd. Er is gekozen voor een yard van 3x3 als beginomgeving, omdat deze gemakkelijk uit te breiden is zowel in breedte als in hoogte. Deze yard geeft de mogelijkheid om de lijst met containers in verschillende vormen in te delen.
Het aantal zeevaartschepen en het aantal containers per schip wordt meegegeven en op basis daarvan wordt de lijst met containers gegenereerd. Dit hoeft niet per se evenveel of meer containers te zijn dan het maximaal aantal plekken. Het model zal de episode stoppen wanneer de lijst met containers leeg is of wanneer de yard vol is.

Methodologie

Opbouw van het model

Voor het toepassen van het PPO-model zijn er een aantal belangrijke factoren die benoemd moeten worden. Zo is het onder andere van belang dat het duidelijk is voor welke action space is gekozen. De action space die aan het model wordt meegegeven is een discrete actie die een x en y coördinaat meegeeft. Deze x en y coördinaat staan voor de plek waar de container wordt geplaatst in de yard, die in dit geval weergegeven is als een matrix. Aan het model wordt ook een observation space meegegeven. Deze observation space bestaat uit het environment waarin wordt gewerkt, die bestaat weer uit twee verschillende onderdelen. Voor een 3x3 yard gaat het om een box van drie bij drie en het nummer van de container die geplaatst gaat worden in die zet. Om ervoor te kunnen zorgen dat het model leert van zijn acties, moet het model fouten of slechte zetten gaan herkennen. Door een reward-functie op te stellen kan er een score per zet worden berekend. Wanneer het model bijvoorbeeld een container plaatst op een plek waar deze niet optimaal staat kan er een negatieve score worden gegeven. Hierdoor zal het model gaan leren om in het vervolg die container niet meer op die plek te plaatsen. Met de som van alle zetten kan er een eindscore worden berekend. Het model zal gaan proberen om elke nieuwe iteratie zijn eindscore te verbeteren. Op die manier traint het model om een betere opstelling van containers te creëren. Om het model te trainen wordt er aangegeven hoeveel timesteps deze moet doorlopen.

De gevonden factoren die van invloed zijn op de efficiëntie van containerterminals zijn in dit geval de plaatsing van de containers en de daarbij gepaarde tijd die de stackers nodig hebben om bij een container te komen. Wanneer de stacker eerst container A moet verplaatsen om bij container B te kunnen is dat niet optimaal, deze handeling kost meer tijd dan wanneer de stacker meteen container B kan pakken. Om de stacker in eerste instantie bij alle gerichte containers te laten komen, zijn er twee factoren die hierin een rol spelen. Aangezien de stackers de containers alleen vanaf de lange zijde kunnen verplaatsen, is het niet handig om containers in te boxen. Met inboxen wordt bedoeld dat er een container geplaatst wordt waardoor een andere container of deze container niet meer bereikbaar is als die container direct moet worden verplaatst. De stacker kan dan dus niet meteen bij de geïnitieerde container, maar heeft daar een extra stap voor nodig (zie Bijlage, Figuur 5). Hierbij wordt in het model alleen gekeken naar de onderste laag containers bij het plaatsen. Daarnaast is het belangrijk dat er geen gaten ontstaan tussen containers bij het plaatsen van de containers. Met de stackers is het namelijk zo dat deze geen container kan pakken of plaatsen wanneer er aan beide lage zijdes van de geïnitieerde container nog containers staan. Het is dus handig om de containers bij plaatsing dicht bij de containers te plaatsen die voor hetzelfde zeevaartschip bestemd zijn. Nu de knelpunten zijn geïdentificeerd kan daar rekening mee worden gehouden bij het creëren van het model. Dat wordt gedaan door in de reward-functie rewards en penalty's toe te kennen waardoor het model gaandeweg leert wat het meest optimale is.

Evaluatie

Na het afronden van de training van het model, is het noodzakelijk om de efficiëntie van de ingedeelde kade te evalueren. Dit wordt handmatig gedaan door het uitvoeren van een visuele inspectie van de ingedeelde kade die het model heeft gegenereerd. Hiervoor is naast het model, een applicatie ontwikkeld die de output van het model in een 3d omgeving kan visualiseren en valideren (zie Bijlage, Figuur 6). Hierdoor kan er worden gezien hoe het model de kade heeft ingedeeld en of er verbeteringen kunnen worden aangebracht. Naast de evaluatie van de indeling van de kade, zal ook het model zelf geëvalueerd worden door middel van de analyse van de leerprestaties van het model over het totaal aantal timesteps, afgebeeld in grafieken. Hierdoor kan er worden bepaald of het model zo optimaal mogelijk presteert.

Een manier om de leerprestaties te analyseren is door te kijken naar de loss functie van het model. Bij de loss functie wordt gekeken naar hoe de actor-critic (Trivedi, 2021) relatie zich heeft ontwikkeld. De critic geeft bij elke actie van de actor een score. Als de critic een penalty geeft zal er een loss plaatsvinden. Door te kijken naar hoe de loss waardes staan tegenover de timesteps die worden genomen, is te zien of het model leert. Bij een minimale loss zal de agent een minimale penalty krijgen. Verder kan er worden gekeken naar de value loss. De value loss geeft aan hoe goed het model is in het beoordelen van de states (Medium, 2021). De value loss zal in het begin omhooggaan tijdens het leren en zal daarna dalen als de beloningen stabiel worden. De value loss kan een maatstaf zijn in het meten van hoe stabiel het model is. Bij een minimale waarde zal het model bijna altijd dezelfde uitkomst bieden.

Complexiteit van het model

Om te concluderen of het model aan het doel voldoet, worden er verschillende situaties uitgewerkt. Allereerst wordt er gekeken naar een kade in een vorm van een matrix van 3x3. Hier kunnen negen containers geplaatst worden. Wanneer dit werkt kan er worden gekeken naar een uitbreiding in de vorm van een matrix van 4x4. Er komt dan wel een extra kolom waarbij er containers ingeboxed kunnen worden. Daar moet rekening mee gehouden worden in de reward-functie. De reward-functie zal dus een kleine aanpassing moeten krijgen. Nadat een 4x4 matrix optimaal werkt kan er worden gekeken naar een 5x5 matrix. Hier komt er nog een extra kolom bij en zal de reward-functie dus weer een kleine aanpassing krijgen.

Nadat het blijkt dat er redelijk makkelijk kan worden uitgebreid in de lengte en breedte, is het idee om dat ook te gaan onderzoeken voor de hoogte. Bij een containerterminal worden de containers immers ook de hoogte in gestapeld. Hier komen wel een aantal andere componenten bij kijken. Tot hiervoor werd er gewerkt in een tweedimensionale omgeving. De observation space wordt nu een driedimensionale omgeving. De yard wordt dan weergegeven als Numpy array. Ook de reward-functie moest weer worden bijgeschaafd. Er moest rekening worden gehouden met het feit dat het model containers op elkaar mocht gaan stapelen.

3. Onderzoeksresultaten

Hieronder zal worden beschreven wat de resultaten zijn van het beste eindmodel, dat een 3x3x3 yard indeelt met containers.

Verzamelde data

Om de resultaten van het model te analyseren, worden er twee aspecten geanalyseerd. Het eerste aspect zijn de voorspellingen en visualisatie van het model. Dit aspect geeft inzicht in hoe de containers in de yard zijn geplaatst. Het tweede aspect is de leerprestaties van het model, die worden gemeten en weergegeven in grafieken door middel van gegevens. Deze gegevens worden verzameld tijdens het leerproces van het model. Hierdoor kan worden geanalyseerd of het model het verwachte leerpatroon heeft gevolgd.

Reward-functie

Bij het maken van de reward-functie is er gekeken naar de volgende punten:

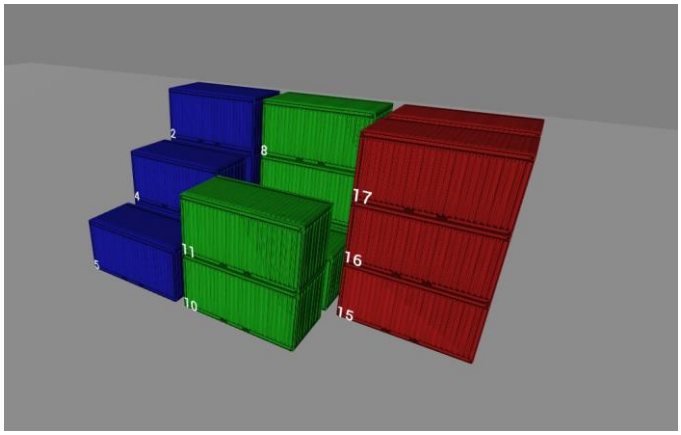
- De container mag niet boven de maximale hoogte komen.
- De container mag niet op een container worden gestapeld die voor een ander zeevaartschip bedoeld is.
- De container mag niet worden ingeboxed

Bij het kijken naar de maximale hoogte wordt alleen een penalty gegeven als de agent een container op een container wil plaatsen die al op de maximale hoogte staat. Hierbij wordt de bovenste container overschreven als dit het geval is. In het geval dat de maximale hoogte nog niet is bereikt wordt er gekeken naar voor welk schip de container eronder bestemd is. Als het voor hetzelfde schip bestemd is, wordt er een beloning toegekend. Als het voor een ander schip bestemd is zal er een penalty worden gegeven. Daarnaast wordt er een beloning gegeven als een container op een lege plek wordt neergezet. Vervolgens, bij het plaatsen van een container op een lege plek, is er een check voor het inboxen van de container. Deze check geeft een penalty als er wordt geconstateerd dat er een container is ingeboxed. Het neerzetten op een lege plek en checken voor inboxen wordt bij elkaar opgeteld en als één waarde teruggegeven als een beloning of penalty.

Visualisatie resultaten

Nadat de agent geleerd heeft, geeft het model een oplossing voor het neerzetten van de containers. De indeling van de containers, die het model maakt na trainen, wordt visueel weergegeven. Deze uitkomst kan vervolgens worden vergeleken met de aanvankelijke situatie en met andere oplossingen om te bepalen hoe efficiënt het model de containers heeft geplaatst. Bovendien kan de uitkomst worden gebruikt als referentie voor toekomstige verbeteringen en vergelijkingen met andere methoden.

In Figuur 1 is te zien dat de containers gesorteerd zijn op kleur, waarbij blauw, groen en rood worden gebruikt als kleuren voor de verschillende zeevaartschepen waarvoor de containers bestemd zijn. Hierdoor is het duidelijk te zien welke containers voor welk zeevaartschip zijn bedoeld. In het figuur is te zien dat de containers zijn gegroepeerd volgens kleurcode, en zo zijn opgesteld dat ze kunnen worden opgepakt door een stacker via de lange zijde. Dit geeft een duidelijke visuele weergave van hoe de containers zijn georganiseerd.



Figuur 1 - Visualisatie output model

Grafieken leerproces modellen

Om te beoordelen of een model effectief leert of vooruitgang boekt tijdens het leerproces, kan men gebruik maken van het visualiseren van de data dat verzameld wordt tijdens het leerproces van het model. In Figuur 2, 3 en 4 worden een aantal visualisaties weergegeven die informatie geven over de prestaties van het model.

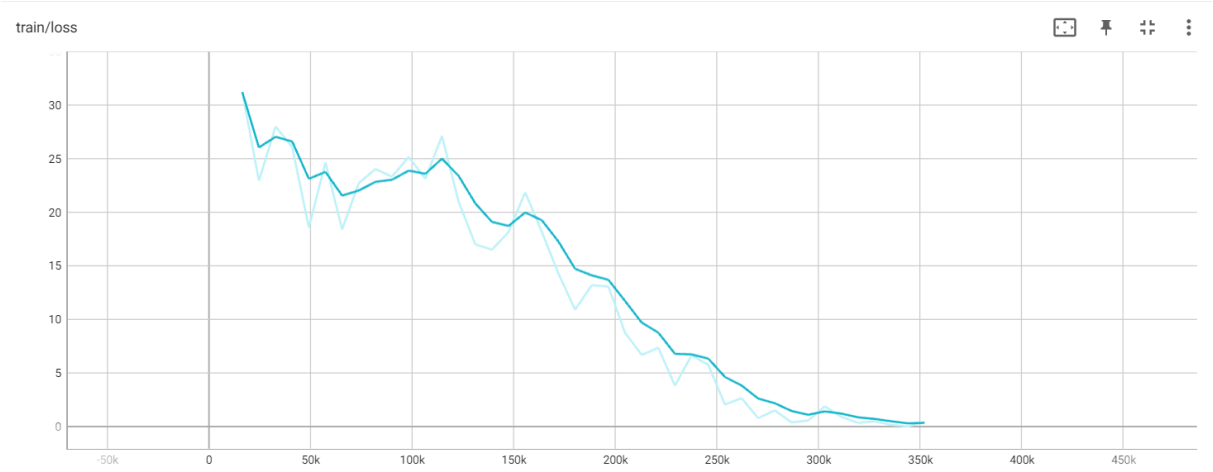
In Figuur 2 worden de scores tijdens het trainen van het model per iteratie getoond, deze resultaten geven aan dat het model effectief is getraind en goed presteert wat betreft de gekozen inputvariabelen.

time/	
fps	1054
iterations	43
time_elapsed	334
total_timesteps	352256
train/	
approx_kl	0.020521311
clip_fraction	0.173
clip_range	0.2
entropy_loss	-0.469
explained_variance	0.985
learning_rate	0.0005
loss	0.456
n_updates	420
policy_gradient_loss	-0.0126
value_loss	1.11

Figuur 2 - Prestaties model

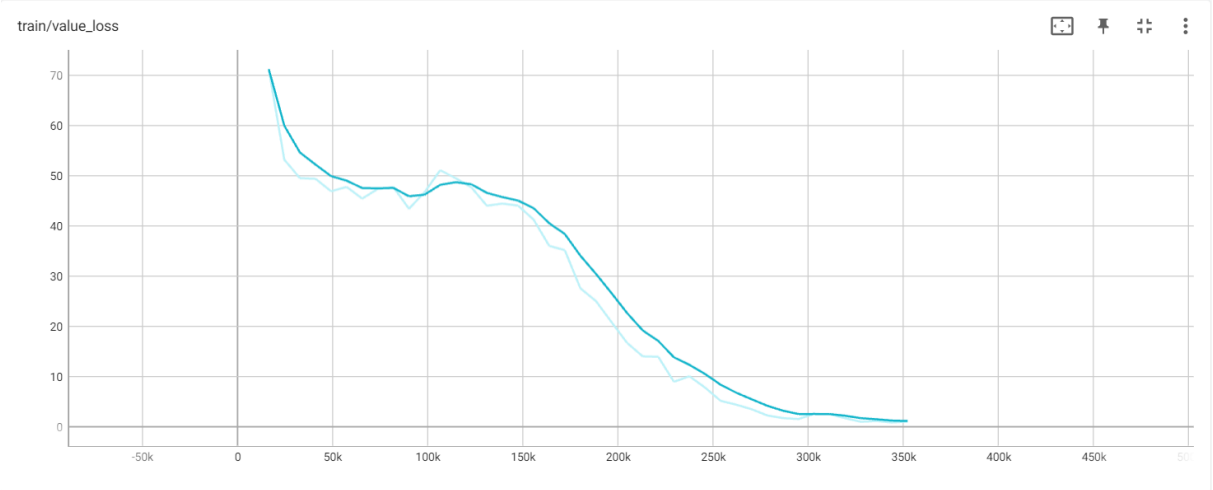
In de Figuren 3 en 4 zijn de waarde van de loss functie en de value loss functie geplot over het aantal timesteps. De doorzichtige lijn is de werkelijke waarde aan, terwijl de dikgedrukte lijn een gladgetrokken versie van de doorzichtige lijn is.

In Figuur 3 is de loss van het model te zien. Door te kijken naar hoe de loss waardes staan tegenover de timesteps die worden genomen, is te zien of het model leert. In dit geval zien we dat het model goed leert.



Figuur 3 - De loss van het model

In Figuur 4 is de value loss van het model te zien, deze beschrijft de stabiliteit van het model over het aantal genomen timesteps.



Figuur 4 - De value loss van het model

Analyse

De prestaties van het PPO-model zijn grondig geanalyseerd om te bepalen hoe effectief het is in het indelen van de containers. Dit is gedaan door middel van een evaluatie van de visualisatie van de voorspellingen en de verzamelde gegevens van het leerproces. In de visualisatie van de geplaatste containers (zie Figuur 1), is te zien dat iedere kleurcode achter elkaar is geplaatst via de lange zijde. In de afgebeelde grafieken wordt opgemerkt dat bij een uiteindelijke timestep van circa 350.000, beide grafieken de nul benaderen en de curve afvlakt.

4. Conclusie

Het doel van dit onderzoeksverslag is het maken van een model dat zo optimaal mogelijk een containerterminal kan indelen en makkelijk uit te breiden is naar een grotere containerterminal. Waarmee er rekening wordt gehouden met een onzekerheid in de binnenkomst van de containers vanaf de binnenvaartschepen. Uit de onderzoeksresultaten blijkt dat er een model is ontwikkeld dat een optimale indeling kan genereren voor een containerterminal van 3x3x3. Er wordt gekeken naar de loss van het gemaakt model, zodra deze loss minimaal is kan er worden gesteld dat het model tot zijn beste reward is gekomen.

Uit Figuur 3 en Figuur 4 blijkt dat de loss en value loss van het model dat is ontwikkeld met ongeveer 350.000 timesteps en in 334 seconden worden geminimaliseerd. De uiteindelijke eindopstelling die dit model genereerd wordt dan handmatig geëvalueerd door te bekijken of de stacker direct bij al de containers van elk zeevaartschip kan. In Figuur 1 is te zien dat dit het geval is. Daar is te zien dat het niet uit maakt of zeevaartschip 1, 2 of 3 als eerste aan komt. De stackers kunnen zonder extra handelingen bij de containers die zijn bestemd voor het zeevaartschip aan de kade. Het doel om een 3x3x3 yard optimaal mogelijk in te kunnen delen is op deze manier bereikt.

Dat het andere doel, het maken van een schaalbaar model, bereikt is, kan worden aangetoond door het feit dat er vanaf een kleine beginsituatie is toegewerkt naar een complexere eindsituatie. Dat is deels gelukt, omdat er is gewerkt vanaf het invullen van een 3x3 yard naar het invullen van een 3x3x3 yard. Deze 3x3x3 yard is dan op de meest optimale manier ingevuld. Echter zijn er wel een aantal kleine aanpassingen gedaan aan het model. Zo moest er bijvoorbeeld een aantal aanpassingen gedaan worden in de reward-functie, action space en observation space toen er van een 2-dimensionale opstelling naar een 3-dimensionale opstelling werd uitgebreid.

5. Discussie en Aanbevelingen

Schaalbaarheid

Een aanbeveling voor verder onderzoek zou zijn om de reward-functie schaalbaar te maken, zonder dat de code aangepast hoeft te worden. Hierdoor zou het mogelijk zijn om de agent te trainen op verschillende vormen, groottes en locaties van yards. In de bestudeerde literatuur over dit specifieke probleem, wordt veelal gepoogd om het aantal bewegingen van de stackers te minimaliseren tijdens het verplaatsen van containers. Dit is ook wat in dit onderzoek bereikt zou moeten zijn door het vinden van de optimale indeling van de yard, zodat de stackers overal snel bij kunnen. Echter wordt in de reward-functie geen rekening gehouden met het aantal zetten die de stackers moeten afleggen. Deze manier van zetten minimaliseren zou een efficiëntere manier kunnen zijn voor het bepalen van de reward, wanneer het probleem groter en complexer wordt.

Wanneer een yard van grootte veranderd of er meer schepen met meer containers binnen komen, wordt het probleem exponentieel moeilijker om op te lossen. Een grotere yard brengt nieuwe uitdagingen met zich mee. Bij een yard van voldoende maat is het vrijwel onmogelijk dat containers niet worden ingeboxed door andere containers. Hiervoor moeten er rekening gehouden worden met de verwachte volgorde van zeeschepen. De observation space zou bijvoorbeeld een verwachte lijst van zeevaartschepen kunnen bevatten.

Verder moet de snelheid van het model geoptimaliseerd worden, aangezien het zo'n 5 minuten duurt om alleen al een lege 3x3x3 te vullen met containers. Door het model slimmer te laten trainen kan het sneller gaan leren. Een voorbeeld om het model slimmer te laten trainen, is door te trainen met meer iteraties waarin er stapsgewijs van een volle yard naar een lege yard wordt getraind. Het model zal dan mogelijk sneller leren om de laatste containers op juiste plaats neer te zetten.

Rekening houden met extra containers

In Figuur 1 is een optimale eindoplossing te zien. Dit komt omdat de stacker nu overal bij kan en er geen containers meer geplaatst hoeven worden in de yard. Echter, wanneer er nu nog meer groene containers binnen zouden komen, is dit eigenlijk geen optimale oplossing. Er is namelijk nog maar 1 plek waar de groene container geplaatst kan worden en dat is bovenop container 11 en de plek achter container 11 is onbereikbaar geworden. In het huidig geformuleerde probleemdomein hoeft hier nog geen rekening mee gehouden te worden. Echter, in het vervolg zal gekeken moeten worden naar het toekennen van een extra penalty voor het plaatsen van containers op deze manier. Zo kan het model ook rekening houden met een optimaal gebruik van de ruimte om rekening te houden met eventueel nieuwe binnenkomende containers.

Validatie Systeem

Om de opstelling die is gegenereerd door het model te evalueren wordt er handmatig gekeken of de stacker direct bij al de containers van elk zeevaartschip kan. Wanneer het probleem complexer wordt is het te ingewikkeld om alles nog handmatig te valideren. Het is om die reden aan te raden om in het vervolg, onderzoek te doen naar een systeem waarmee de opstelling van containers geëvalueerd kan worden.

Reward-functie

In de bestudeerde literatuur over dit specifieke probleem, wordt veelal gepoogd om het aantal bewegingen van de stackers te minimaliseren tijdens het verplaatsen van containers. Dit is ook wat in dit onderzoek bereikt zou moeten zijn door het vinden van de optimale indeling van de yard, zodat de stackers overal snel bij kunnen. Echter wordt in de reward-functie geen rekening gehouden met het aantal zetten die de stackers moeten afleggen. Deze manier van zetten minimaliseren zou een efficiëntere manier kunnen zijn voor het bepalen van de reward, wanneer het probleem groter en complexer wordt.

Aanbeveling visualisatie

Een aanbeveling om de visualisatie applicatie verder uit te breiden zou zijn om een stapsgewijze evaluatie toe te voegen. Dit betekent dat er na elke verplaatsing van een container, een evaluatie plaats zou vinden. Op deze manier kan de gebruiker de verplaatsing van de container real-time volgen en direct zien of de container op de juiste plek staat. Dit is een significante verbetering ten opzichte van de huidige manier van werken, waarbij voor elke zet een nieuw scenario aangemaakt moet worden.

Daarnaast kan de stapsgewijze evaluatie ook helpen bij het identificeren van eventuele problemen of fouten die zich voordoen tijdens het verplaatsen van de container. Bijvoorbeeld, als een container door een andere container heen wordt verplaatst, kan dit direct worden opgemerkt en worden opgelost. Dit kan de efficiëntie van het proces verhogen door eerder het model te straffen voor fouten. In combinatie met een gebruiksvriendelijke interface, kan de stapsgewijze evaluatie een belangrijke bijdrage leveren aan de kwaliteit van het model.

Literatuurlijst

- Hu, H., Yang, X., Xiao, S., & Wang, F. (2023). Anti-conflict AGV Path Planning in Automated Container Terminals Based on Multi-agent Reinforcement Learning. *International Journal of Production Research, 61*(1), 65-80. doi: 10.1080/00207543.2021.1998695

- Hu, X., Yang, Z., Zeng, Q. (2012) A Method Integrating Simulation and Reinforcement Learning for Operation Scheduling in Container Terminals. *Transport, 26*(4), 383-393. doi: 10.3846/16484142.2011.638022

- Jiang, T., Zeng, B., Wang, Y., & Yan, W. (2021) A New Heuristic Reinforcement Learning for Container Relocation Problem. *Journal of Physics: Conference Series, 1873*(1). 012050. doi: 10.1088/1742-6596/1873/1/012050

- Kefi, M., Korbaa, O., Ghedira, K., & Yim, P. (2007). Heuristic-based model for container stacking problem. In *19th International Conference on Production Research-ICPR* (Vol. 7).

- Krishna, V., Sudhir, Y. (2020). *Comparison of Reinforcement Learning Algorithms* [Powerpoint-slides]. Departure of Computere Science and Engeneering, University at Buffalo. Geraadpleegd op 28 november 2022, van https://cse.buffalo.edu/~avereshc/rl_fall20/

- Medium. (2021, 7 december). *Understanding PPO plots in TenserBoard.* Geraadpleegd op 21 januari 2023, van https://medium.com/aureliantactics/understanding-ppo-plots-in-tensorboard-cbc3199b9ba2

- Renotte, N. (2021, 6 juni). *Reinforcement Learning in 3 hours | Full Course Using Python* [Video]. YouTube. Geraadpleegd op 21 november 2022, van https://www.youtube.com/watch?v=Mut_u40Sqz4&t

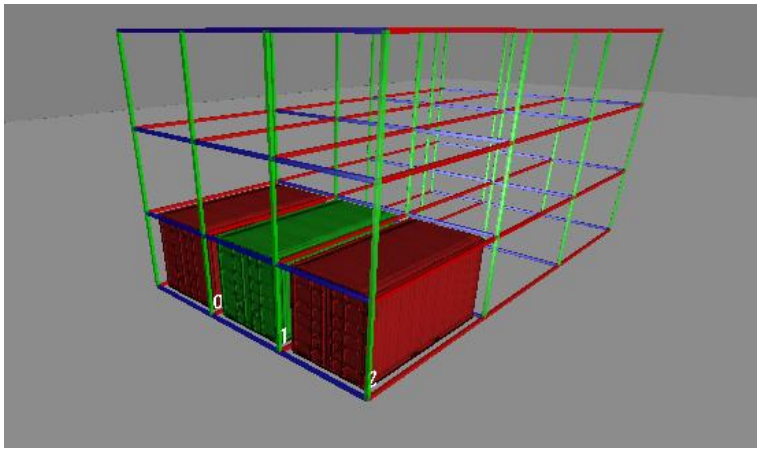
- Ries, J., González-Ramírez, R. G., Miranda, P.(2014). A Fuzzy Logic Model fort he Container Stacking Problem at Container Terminals. *International Conference on Computational Logistics,* 93-111. doi: 10.1007/978-3-319-11421-7_7

- Salido, M. A., Sapena, O., & Barber, F. (2009). An artificial intelligence planning tool for the container stacking problem. *2009 IEEE Conference on Emerging Technologies & Factory Automation*, 1-4. doi: 10.1109/ETFA.2009.5347007.

- Trivedi, C. (2021, 11 december). *Proximal Policy Optimization Tutorial (Part 2/2: GAE and PPO loss)*. Medium. Geraadpleegd op 21 januari 2023, van https://towardsdatascience.com/proximal-policy-optimization-tutorial-part-2-2-gae-and-ppo-loss-fe1b3c5549e8

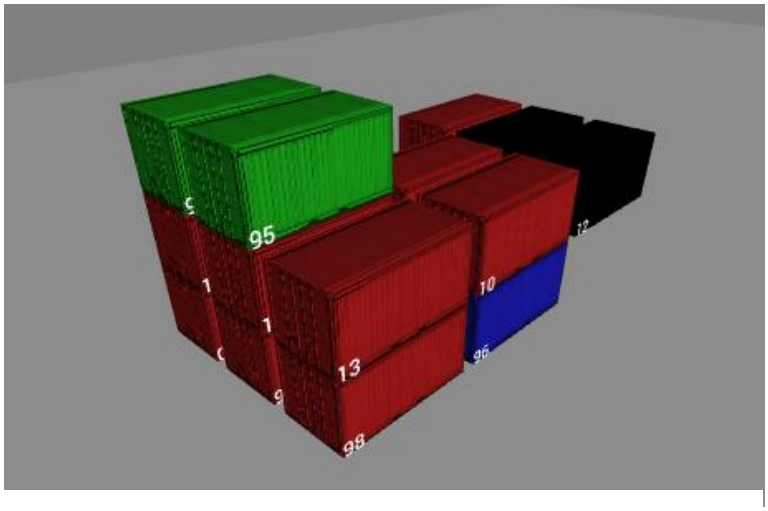
Bijlage

Een groene container wordt van beide lange zijden ingeboxed door rode containers en is hierdoor niet in een zet bereikbaar.



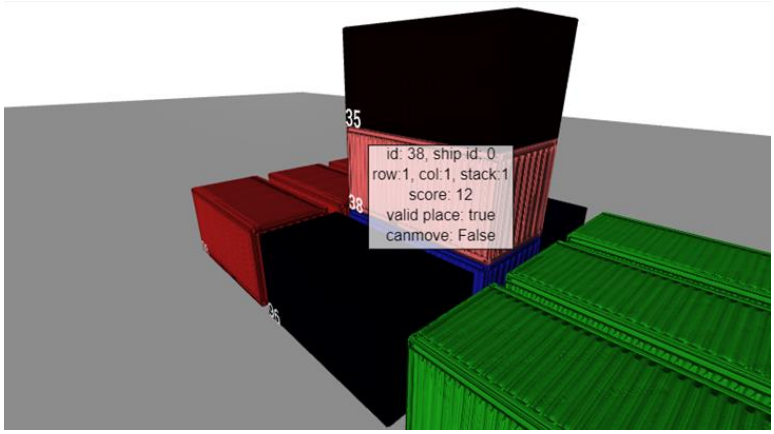
Figuur 5 - Inboxen

De output van het model wordt geëvalueerd om te bepalen of een container correct is geplaatst. Als de container op een onjuiste manier is geplaatst, wordt dit aangegeven door de container in het zwart weer te geven. In dit specifieke voorbeeld worden er container in de lucht geplaatst, dat aangeeft dat de containers niet correct geplaatst zijn.



Figuur 6 - Validatie

Container met nummer 38 wordt geblokkeerd door de containers weergegeven met een zwartte kleur.



Figuur 7 – Geblokkeerde container