

Werkplan Masterproef

Titel	Interpretable deep reinforcement learning policies
Naam Student	Xander De Visch
Email	Xander.DeVisch@UGent.be
Bedrijf/ Onderzoeksgroep	IDLab imec
Promotoren	Promotor 1: prof. dr. ir. Pieter Simoens Promotor 2: dr. Sam Leroux
Begeleiders	Mattijs Baert

Opsplitsing per semester:

*Duid aan hoeveel studiepunten (vakken) je in elk semester volgt (behalve de masterproef). Verdeel ook de **18** studiepunten van de masterproef over de twee semesters. (Eén studiepunt komt overeen met 25-30 uren)*

	Semester 1	Semester 2
# studiepunten vakken	24	18
# studiepunten masterproef	6	12

Bestaande situatie en probleemstelling

Geef duidelijk aan wat er nu gebruikt wordt, en waar het probleem zich situeert (minstens een halve pagina).

Deep reinforcement learning (DRL) is een vorm van machinaal leren en artificiële intelligentie waarbij machines kunnen leren van hun acties, zeer gelijkaardig aan hoe mensen leren van hun ervaringen. De agent observeert de huidige status van een omgeving, kiest een actie en ontvangt een beloningssignaal dat de kwaliteit van de gekozen actie weergeeft. Het doel is om een beleid te vinden dat de beloning maximaliseert.

DRL maakt gebruik van *deep neural networks* (DNNs). DNNs hebben de laatste jaren zeer sterke resultaten behaald en bleken specifiek in het gebied van computervisie superieur tegenover andere algoritmen in het gebied van machinaal leren.

Het probleem echter is de *black-box* natuur van DNNs. Door de hoge dimensionele feature-ruimte en de complexe architectuur is het duidelijk dat het uitrollen van DRL-instanties in zeer kritieke omgevingen zoals zelfrijdende auto's, robotica in dienstsectoren, gezondheidszorg, etc. een gevoelige zaak kan zijn en delicate consequenties met zich mee kan brengen.

Een ander essentieel probleem is de verklaarbaarheid. Voor systemen die onderhouden worden door mensen. De onderhouders vertrouwen op deze systemen en moeten inzicht krijgen op faalgevallen. In het geval van *policy errors* is het cruciaal dat de oorsprong van het probleem a posteriori gekend is.

Doelstelling van de masterproef

Beschrijf de concrete opdracht voor de masterproef – dit moet meer gedetailleerd zijn dan in het voorstel op Plato (geen copy/paste – minstens een halve pagina).

De doelstelling van de masterproef is om *interpretable machine learning*-modellen te onderzoeken en uit te werken. *Interpretable machine learning*-technieken kunnen een antwoord leveren op de problemen zoals de transparantie, verklaarbaarheid en interpreteerbaarheid van de DRL-systemen met intrinsieke *black-box* structuur. *Interpretable machine learning*-technieken zoals bijvoorbeeld beslissingsbomen kunnen echter geen hoog dimensionale data zoals afbeeldingen behandelen.

In deze thesis moet er een afweging gemaakt worden tussen de predictieve performantie van de DRL-technieken en de interpreteerbaarheid van de beslissingsbomen. De combinatie van *interpretable machine learning*-modellen en neurale netwerken tot een interpretable DRL-algoritme ligt aan de basis van dit onderzoek.

De Neural Prototype Tree, kortweg ProtoTree heeft het representatief vermogen en bevat een ingebouwde binaire boomstructuur. Doorheen de masterproef zal deze verder bestudeerd worden met als einddoel een werkende implementatie van dit model te verkrijgen. Dit model zal uiteindelijk op verschillende *reinforcement learning benchmark* domeinen (waaronder de klassieke Atari omgevingen) van de gym library aangeboden door OpenAi, getest worden.

Planning en mijlpalen

Definieer enkele duidelijk afgebakende taken. Vermeld voor elke taak het aantal weken dat je hiervoor nodig zal hebben en een concrete deadline. Geef een concrete omschrijving van het doel en wat er op het einde van de taak afgeleverd zal worden. Een typische taak neemt 3 tot 5 weken in beslag. Taken kunnen overlappen (bv. Schrijven van de scriptie en uitvoeren van aanvullende experimenten). Hieronder een aantal voorbeelden.

Taak 1	4 weken	Deadline: 20 oktober 2022	Literatuurstudie DQN en technologieverkenning pytorch en gym
Inhoud			
Literatuurstudie DRL pytorch framework en gym library leren kennen en eerste implementatie DQN			
Belangrijkste resultaten, deliverables of inzichten na deze fase:			
Literatuurstudie werking DQN, probleemsituatie black-box in scriptie neerschrijven. Werkend DRL-programma in pytorch maken dat probleem "Frozen Lake" kan oplossen. Resultaten van programma in weights and biases tool met elkaar vergelijken. Meerdere episodes vergelijken en verschillende implementaties vergelijken.			

Taak 2	5 weken	Deadline: 24 november 2022	ProtoTree leren kennen en werkende krijgen in pytorch
Inhoud			
Literatuurstudie ProtoType tree, interpretable neural network trainen met supervised code, policy uitgedrukt als prototype tree leren doormiddel van supervised learning van een pre-trained model			
Belangrijkste resultaten, deliverables of inzichten na deze fase:			
ProtoType tree onderzoeken aan de hand van papers, complete werking en mogelijkheden onderzoeken en neerschrijven in scriptie. Resultaten van supervised code en supervised learning plotten en beoordelen.			

Taak 3	5 weken	<i>Deadline:</i> 24 februari 2022	Eerste vorm van interpretable DRL in actie zetten
--------	---------	--------------------------------------	--

Inhoud

ProtoType tree uit taak 2 proberen verwerken in uitgewerkt DQN voorbeeld van taak 1 en uittesten op eenvoudig discreet environment.

Belangrijkste resultaten, deliverables of inzichten na deze fase:

Interpretable deep reinforcement-model werkende krijgen, resultaat vergelijken met resultaat uit Taak 1 aan de hand van weight and biases.

Resultaat en ondervindingen neerschrijven in scriptie.

Taak 4	7 weken	<i>Deadline:</i> 31 maart 2022	Schrijven aan scriptie
--------	---------	-----------------------------------	-------------------------------

Inhoud

scriptie 1e versie (95%) voorleggen aan promotoren/begeleiders

Belangrijkste resultaten, deliverables of inzichten na deze fase:

Een opgemaakte scriptie volgens de template voorleggen die voldoet aan de opgelegde voorwaarden.

Taak 5	7 weken	<i>Deadline:</i> 31 maart 2022	Interpretable DRL-model met ProtoTree uitwerken op complexe omgeving
--------	---------	-----------------------------------	---

Inhoud

Overschakelen van eenvoudig discrete omgeving naar complexere omgevingen zoals in Atari videospellen of Super Mario.

Belangrijkste resultaten, deliverables of inzichten na deze fase:

Einddoel van project waarbij beleid van complexe omgeving aan de hand van de ProtoTree in een deep reinforcement-model kan weergegeven worden.

Taak 6	2 weken	<i>Deadline:</i> 27 april 2022	Eerste versie van presentatie maken
--------	---------	-----------------------------------	--

Inhoud

Eerste versie van presentatie in orde brengen om te presenteren voor eerstejaarsstudenten/ derdejaarsstudenten.

Belangrijkste resultaten, deliverables of inzichten na deze fase:

Duidelijke presentatie waarbij de situatie, het probleem en de voorgestelde oplossing helder wordt uitgelegd. Aan de hand van feedback van studenten en professoren derde jaar inzichten en werkpunten verkrijgen.

Taak 7	3 weken	Deadline: 27 april 2022	Scriptie en eindpresentatie afwerken
Inhoud			
De finale versie van scriptie en eindpresentatie in orde brengen			
Belangrijkste resultaten, deliverables of inzichten na deze fase:			
Presentatie aanpassen naar de werkpunten verkregen uit taak 6. Scriptie afwerken en maken dat alles voldoet aan de criteria.			

Taak 8	3 weken	Deadline: 28 juni 2022	Vorbereiden op openbare verdediging
Inhoud			
Verdediging van masterproef volledig voorbereiden			
Belangrijkste resultaten, deliverables of inzichten na deze fase:			
Presentatietechnieken en communicatiemodules herbekeken, voorbereid op 'confrontatie' met het uitgebreid voorstel. Beamer vooraf uitgetest.			

Contactmomenten

Vermeld duidelijk hoe de tussentijdse communicatie met de begeleiders en promotoren zal verlopen. Geef bijvoorbeeld aan of er een wekelijkse/ tweewekelijkse/ maandelijkse rapportering zal zijn en of die fysiek, digitaal of via e-mail zal doorgaan.

De tussentijdse communicatiemomenten verlopen wekelijks on campus en gaan door in I Gent in Zwijnaarde (Campus Ardoyen). Indien het niet mogelijk is fysiek af te spreken, verloopt het via Microsoft Teams.

In de wekelijkse rapportering wordt er gekeken wat er de afgelopen week verwezenlijkt is, welke problemen er terecht gekomen zijn, en wat er in de volgende week behandelt zal worden. De vragen van afgelopen week worden beantwoord.

Gantt chart:

	Taak1	Taak2	Taak3	Taak4	Taak5	Taak6	Taak7	Taak8
Week 1								
Week 2	X							
Week 3	X							
Week 4	X							
Week 5	X							
Week 6		X						
Week 7		X						
Week 8		X						
Week 9		X						
Week 10		X						
Week 11			X					
Week 12			X					
Week 13			X					
Kerstvakantie								
Examenperiode januari								
Week 1			X	X				
Week 2			X	X				
Week 3				X	X			
Week 4				X	X			
Week 5				X	X			
Week 6				X	X			
Week 7				X	X			
Paasvakantie								
Week 8					X	X		
Week 9					X	X		
Week 10							X	
Week 11							X	
Week 12							X	
Week 13								X
Week 14								X
Week 15								X