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 feauture ruimte en de complexe architectuur is het duidelijk dat het uitrollen van DRL instanties in zeer kritieke omgevingen zoals zelf rijdende auto’s, robotica in diensectoren, 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 sytemen 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 transparantheid, verklaarbaarheid en interpreteerbaarheid van de DRL systemen met intrinsieke *black-box* structuur. Interpretable machine learning technieken zoals bijvoorbeeld beslissingsbomen kunnen echter geen hoog dimensionele 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 | 3 weken | *Deadline:*  30 oktober 2022 | **Literatuurstudie en technologieverkenning** |
| **Inhoud**  Wetenschappelijke artikelen opzoeken en samenvatten over de onderwerpen X, Y en Z. | | | |
| **Belangrijkste resultaten, deliverables of inzichten na deze fase:**  Eerste versie van de literatuurstudie, neergeschreven in de scriptie  Onderzoeksvragen duidelijker afgebakend en doelstelling voor einde eerste semester geconcretiseerd.  Opsomming van de belangrijkste bestaande technieken om XYZ op te lossen. | | | |

|  |  |  |  |
| --- | --- | --- | --- |
| Taak 2 | 3 weken | *Deadline:*  21 december 2022 | **Implementatie baselines** |
| **Inhoud**  Op basis van de resultaten van de literatuurstudie worden er 3 relevante baseline technieken geselecteerd en uitgetest op de bestaande dataset. | | | |
| **Belangrijkste resultaten, deliverables of inzichten na deze fase:**  Voor elke techniek wordt de performantie en uitvoeringstijd gemeten. Deze resultaten worden in een tabel en grafiek weergegeven om de trade-off tussen beiden te analyseren. | | | |

...

|  |  |  |  |
| --- | --- | --- | --- |
| Taak 5 | 2 weken | *Deadline:*  30 maart 2023 | **Scriptie – eerste 25 bladzijden** |
| **Inhoud**  Starten met de scriptie: structuur, inleiding, literatuurstudie | | | |
| **Belangrijkste resultaten, deliverables of inzichten na deze fase:**  Juiste structuur voor de scriptie: wat wordt besproken? | | | |

|  |  |  |  |
| --- | --- | --- | --- |
| Taak 1 | 4 weken | *Deadline:*  20 oktober 2022 | **Literatuurstudie DQN en technologieverkenning pytorch en gym** |
| **Inhoud**  Literatuurstudie Deep reinforcement learning, 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 neergeschreven.  Werkend Deep Reinforcement Learning programma in pytorch maken dat cartpole probleem kan oplossen.  Resultaten van programma in weights and biases tool met elkaar vergelijken. Meerdere episodes vergelijken en verschillende implementaties vergelijken. | | | |

|  |  |  |  |
| --- | --- | --- | --- |
| Taak 2 | 3 weken | *Deadline:*  10 november 2022 | **ProtoTree leren kennen en werkende krijgen in pytorch** |
| **Inhoud**  Literatuurstudie ProtoType tree, interpretable neural netwerk 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 | *Deadline:*  12 december 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 (tool).  Resultaat en ondervindingen neerschrijven in scriptie. | | | |

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 IGent 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.

Grantt 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 |  |  |  |  |  |  |  |  |
| Week 2 |  |  |  |  |  |  |  |  |
| Week 3 |  |  |  |  |  |  |  |  |
| Week 4 |  |  |  |  |  |  |  |  |
| Week 5 |  |  |  |  |  |  |  |  |
| Week 6 |  |  |  |  |  |  |  |  |
| Week 7 |  |  |  |  |  |  |  |  |
| Week 8 |  |  |  |  |  |  |  |  |
| Week 9 |  |  |  |  |  |  |  |  |
| Week 10 |  |  |  |  |  |  |  |  |
| Week 11 |  |  |  |  |  |  |  |  |
| Week 12 |  |  |  |  |  |  |  |  |
| Week 13 |  |  |  |  |  |  |  |  |
| Week 14 |  |  |  |  |  |  |  |  |
| Week 15 |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |