WETENSCHAPPEN &
BIO-INGENIEURSWETENSCHAPPEN

De Onderzoeksgroep
Artificieel Intelligentie Lab

nodigt U graag uit op de openbare verdediging van het proefschrift van

Denis Steckelmacher

ter behaling van de graad van Doctor in de Wetenschappen

Titel van het proefschrift:

Modelvrij bekrachtigingsleren voor de echte wereld

Promotor:
Prof. dr. Ann Nowé

De verdediging heeft plaats op

Vrijdag 6 november 2020 om 17u00

De verdediging kan via een livestream gevolgd worden: https://youtu.be/UGRDV6UfUbY

Samenstelling van de jury

Prof. dr. Bernard Manderick (VUB, voorzitter)
Prof. dr. Elisa Gonzalez Boix (VUB, secretaris)
Prof. dr. ir. Bram Vanderborght (VUB)
Prof. dr. Manuela Veloso (Machine Learning Department, School of Computer Science, Carnegie-Mellon University, US)
Prof. dr. Robert Babuska (Department of Cognitive Robotics, Faculty of Mechanical, Maritime and Materials Engineering, Delft University of Technology, NL)

Curriculum vitae

Denis Steckelmacher behaalde in 2016 zijn master diploma Computerwetenschappen aan de Université Libre de Bruxelles en startte zijn doctoraat in 2016 aan de VUB. Zijn onderzoek is gericht op voorbeeldefficiëntie voor bekrachtigingsleren, toegepast op robots. Hij heeft ook ervaring met digitale elektronica en high-performance computing. Hij is co-auteur van 9 peer reviewede artikels in internationale conferenties, presenteerde zijn werk op meerdere internationale events en won de prijs voor de beste demo op BNAIC 2017 en 2019. Hij gaf les aan de VUB, EHB Brussel, een winterschool in Delft en een zomerschool in Nieuwpoort.

Abstract van het doctoraatsonderzoek

Met bekrachtigingsleren kan een agent leren hoe een taak te volbrengen. De agent observeert toestanden, voert acties uit en ontvangt positieve of negatieve beloningen na elke actie. Het doel van de agent is om te leren welke actie in welke situatie uit te voeren om de totale beloningen te maximaliseren. Eén van de belangrijke uitdagingen voor de verdere toepassing van bekrachtigingsleren is net dat het moet leren om steeds beter te worden uit ervaring. Dit betekent dat de agent aanvankelijk suboptimaal is en maar geleidelijk aan beter wordt. In de echte wereld, vooral wanneer de agent geen toegang heeft tot een simulator (bv. een fysieke robot die met mensen interageert), consumeert elke suboptimale actie waardevolle tijd en middelen.

Ons onderzoek helpt de agent via *sample-efficiency* (voorbeeld-efficiëntie). Sample-efficiency is een maat die uitdrukt hoe snel, met hoeveel acties een bekrachtigingslerende agent een taak kan leren doen. We beschrijven nieuwe algoritmes die zeer hoge sample-efficiency hebben, dus in staat zijn om snel te leren a.d.h.v. weinig acties. De agent wordt snel “goed genoeg” en voert slechts een beperkt aantal suboptimale acties uit tijdens de leerfase. Doordat een sample-efficiënte agent leert met weinig interacties, kan hij ook met weinig middelen (elektriciteit, supervisie, feedback van een eindgebruiker) getraind worden.

Onze belangrijkste bijdrage is Bootstrapped Dual Policy Iteration (BDPI), een nieuw modelvrij en sample-efficiënt algoritme voor bekrachtigingsleren. We tonen aan dat BDPI uitdagende echte wereld taken kan leren, door een gemotoriseerde rolstoel te laten leren hoe die naar een bepaalde plaats moet rijden in een ongestructureerde omgeving, zoals een rommelig kantoor. BDPI heeft geen simulator, geen vooropleiding en geen voorbereidingen in het kantoor nodig. BDPI gebruikt simpele sensoren (webcams), en kan de taak leren in minder dan een uur. Hiermee demonstreren we dat bekrachtigingsleren de mogelijkheid heeft om complexe taken te leren in de echte wereld.