

LEARN4T - Testen met behulp van AI bij de NS loont!



Artificial intelligence kent sinds enkele jaren een nieuwe veelbelovende benadering: reinforcement learning. Het is bekend geworden door succesvolle toepassingen met spelletjes als GO, maar kan het ook in IT-development succesvol worden toegepast? Bij de NS heeft het Test Competence Centre samen met Improve Quality Services geëxperimenteerd om een antwoord te krijgen op die vraag. In een project is Reinforcement Learning gebruikt om e-learnings te testen om daarmee saai testwerk te vervangen door AI.

Het testen van e-learnings is iets wat in het begin nog wel interessant is, maar al snel vervelend wordt.

Wat is Reinforcement Learning?



Anders dan bij andere benaderingen binnen machine learning, heb je bij reinforcement learning geen data nodig om te kunnen beginnen. De data worden al doende aangemaakt. Het model onderneemt acties en krijgt daar al dan niet een reinforcement, oftewel een beloning, voor. Het model gaat op zoek naar steeds meer beloningen, en 'denkt' slim na hoe deze beloningen het meest effectief binnengehaald kunnen worden. Het leert dus van de beloningen en ontwikkelt de meest efficiënte strategie voor een zo groot mogelijke beloning. Deze vorm van machine learning is ook uitermate geschikt in het oplossen van een rubiks cube, het leren lopen van een robot en het spelen van een spel als Go.

Wat testen we?

Het system under test (SUT) is het e-learning platform van NS. De organisatie kent een grote variatie aan e-learnings en elke week komen er gemiddeld 5 bij. Denk hierbij aan online training voor NS medewerkers in omgang met klanten en werkinstructies voor het gebruik van bepaalde software en hardware onderhoud. Deze e-learnings worden nu handmatig getest met de focus voornamelijk op technische zaken. Meestal opent een tester de e-learning op het nieuwe platform. Daarna wordt er door de e-learning geklikt, filmpjes en geluid wordt gestart en wordt er gekeken of de content goed geladen wordt en de filmpjes en geluid kunnen worden afgespeeld. Dit kost gemiddeld een uur per e-learning. Het testen van e-learnings is iets wat in het begin nog wel interessant is, maar al snel vervelend wordt.

Om een model te trainen met reinforcement learning moeten we acties kunnen uitvoeren en observeren of er een beloning is verdiend.



Wat kost het en wat levert het op?

Veel mensen hebben het idee dat bij machine learning zware machines met veel rekenkracht nodig zijn. Bij reinforcement learning valt dat mee. De bottleneck in ons experiment was de snelheid van de e-learnings. Een gemiddelde laptop kan het inleren van het model tijdens het laden van pagina's en het klikken door een webapp prima bijbenen. De kosten voor het draaien zijn dus vrij laag. Op dit moment wordt het model elke dag bijgeleerd door een aantal keer per dag een willekeurige e-learning te pakken en een uur lang te leren. In de praktijk zien we dat een nieuwe e-learning kan worden doorlopen in minder dan 2 uur. Dat is langzamer dan een handmatige test die gemiddeld een uur in beslag neemt, maar het kan continue op de achtergrond worden uitgevoerd. De



404
Page not found

No image
available

rapportage die er uit komt heeft af en toe analyse nodig, maar met het filteren van fouten wordt deze analyse steeds beter. De manuele testtijd is teruggebracht van iets meer dan 1 uur kijken naar een kwartier gericht zoeken en dit kan verder naar beneden worden gebracht. Wel is het zo dat het algoritme op dit moment niet alle e-learnings aan kan. Zo worden er vragen gemaakt waar specifieke tekstuele input wordt verwacht. Dit soort acties worden op dit moment nog niet ondersteund, maar zijn wel onderwerp van nieuwe experimenten.

Conclusie en hoe verder?

De conclusie die we kunnen trekken is dat het mogelijk is om machinelearning of AI in te zetten in de praktijk. Het gebruik van reinforcement learning kent een vrij korte leercurve en is door veel ontwikkelaars direct toe te passen. Het stelt daarnaast niet veel eisen aan de hardware. De fouten die gevonden worden zijn voornamelijk technische fouten. Filmpjes die niet afspelen, plaatjes die niet worden geladen en scripts die niet goed werken met andere versies. Deze gevonden fouten zijn wel verreweg de grootste risico's bij de e-learnings. Wel is het zo dat er in dit onderzoek een significante tijd is besteed aan het uitzoeken van de juiste parameters en algoritmes.

Samen met de NS zijn we nu bezig om complexere acties te kunnen ondersteunen. Zo willen we invoer toevoegen aan de lijst van acties, al dan niet bepaald door data vanuit productie. Op die manier willen we een LEARN4T 2.0 maken die ook in staat is om low-code apps te testen. De focus ligt vooral in het uitvoeren van complexere acties die bijvoorbeeld invoervelden meebrengen. De eerste resultaten zijn bemoedigend, maar we hebben hier nog wel een lange weg te gaan.



Pieter Withaar is innovatiemanager, consultant en trainer bij Improve Quality Services en onder andere werkzaam bij NS.

pieter.withaar@improveqs.nl