

Intelligent Agents:

een moderne vorm van AI

Prof. Dr. J.-J. Ch. Meyer,

Hoogleraar aan het Instituut voor Informatica en Informatiekunde aan de UU

In de hedendaagse AI is een belangrijke rol weggelegd voor autonome, proactieve en communicerende stukjes software die observeren wat er in hun omgeving gebeurt en daarop kunnen reageren: *Intelligent Agents*. Dergelijke entiteiten zitten volgens bepaalde architecturen in elkaar en worden in al dan niet specifieke programmeertalen verwezenlijkt. Alléén of in een samengesteld systeem zijn ze nuttig in talloze toepassingen, van de geneeskunde tot aan de ruimtevaart, in de entertainment-industrie of in de productie.

In dit korte artikel zal ik een (subjectieve) schets geven van het relatief nieuwe terrein van de agenttechnologie. Alhoewel de artificiële intelligentie (AI) al weer een halve eeuw bestaat als aparte discipline, met allerlei subdisciplines zoals kennisrepresentatie, redeneren, leren, 'vision', etc., is men vrij recentelijk tot het inzicht gekomen dat een integrale aanpak nodig is om intelligente systemen te maken. Het centrale concept hierbij is dat van een (*intelligent*) agent. Een agent observeert, redeneert en handelt (letterlijk is een agent iets dat handelt, van het Latijnse woord 'agere' = handelen) in een omgeving, waarbij in principe alle aspecten van (artificiële) intelligentie een rol spelen. Men zou de agenttechnologie kunnen opvatten als een nieuwe poging om het ideaal van de klassieke (m.n. symbolische) AI (*Good Old-Fashioned AI – GOFAI*) te verwezenlijken. Het is zelfs zo dat Russell & Norvig in hun fameuze leerboek 'Artificial Intelligence, A Modern Approach' (op de voorflap "The Intelligent Agent Book" genoemd!) op bladzijde 1 zeggen: 'AI is the study of agents that exist in an environment and perceive and act'. Een zeer interessante ontwikkeling is verder dat het agentgeoriënteerde paradigma verwant is aan het objectgeoriënteerde paradigma in de software engineering, zodat het een interessante mix van onderzoekers aantrekt: klassieke AI-ers, toegepaste logici en cognitiewetenschappers, maar ook software engineers, en vanwege de 'sociale' aspecten waarover ik het later nog zal hebben ook sociale wetenschappers. Ook zijn er allerlei links met andere disciplines zoals de wiskundige besliskunde en speltheorie.

Wat zijn nu intelligent agents? Er is nogal wat controverse over het begrip agent. Ik houd de volgende definitie aan: een stuk software dat een bepaalde vorm van autonoom gedrag vertoont. Vaak worden hierbij de volgende begrippen geassocieerd: reactief, proactief en sociaal. Hiermee wordt bedoeld dat de agent adequaat reageert op een immer veranderende omgeving, dat deze doelgericht is en 'zelf' initiatieven neemt en dat deze samenwerkt met andere agents.

Ik geef toe: dit is nog steeds tamelijk vaag. In veel agent-literatuur wordt ervan uitgegaan dat agents hun autonome gedrag realiseren middels zogenaamde mentale toestanden en attitudes, zoals het omgaan met kennis, geloof, wensen, doelen, intenties, *commitments*, etc. Het huidige agentonderzoek komt dan ook voort uit filosofisch werk over hoe de mens handelt op grond van zijn mentale attitudes. De filosoof Michael Bratman betoogde in het bijzonder het belang van intenties voor het

doelgericht handelen van mensen. Intenties zijn als het ware doelen waarop 'gecommit' wordt (waar men voor gaat, zo gezegd) en die men niet opgeeft voordat men ze heeft bereikt of voordat men inziet dat ze niet bereikt kunnen worden. Dit heeft te maken met zogenaamde *commitment*-strategieën - over wanneer intenties worden opgegeven - die zijn bestudeerd in de literatuur. In eerste instantie hebben onderzoekers zich

beziggehouden met de vraag hoe dit te beschrijven met logische theorieën. Zeer bekend is de zogenaamde *BDI* (*Belief-Desire-Intention*)-logica, een logisch formalisme voorgesteld door de Australische onderzoekers Rao & Georgeff om agents en hun gedrag (bijvoorbeeld hun *commitment*-strategieën) te specificeren. Maar men is ook vrij snel aan de slag gegaan met de vraag hoe dit soort agents kunstmatig zouden kunnen worden gerealiseerd. In feite heeft Bratman zelf met de hulp van informatici/AI-onderzoekers al een eerste architectuur ontworpen voor zulke kunstmatige agents. Het is natuurlijk maar zeer de vraag of op artificiële agents wel dezelfde mentale concepten (beliefs, desires, intentions) van toepassing zijn als op mensen. Dit hangt ook erg van iemands filosofie van AI af ('sterke' of 'zwakke' AI). De meer 'nuchtere' onderzoeker zal echter nog steeds de genoemde concepten kunnen gebruiken als een bruikbare en nuttige metafoor!

Inmiddels zijn er een aantal agent-architecturen voorgesteld, die aangeven hoe de agent op basis van zijn mentale toestand handelt en hoe deze toestand over de tijd verandert door zijn handelingen. Het proces dat de agent uitvoert om tot een keuze van uit te voeren handelingen te komen (en meer algemeen ook hoe deze kiest welke motivationele attitudes zoals wensen, doelen en intenties hierbij een rol spelen) heet de deliberatiecyclus van de agent. Omdat een agent (afhankelijk van de toepassing) ook af en toe in een soort reflex moet kunnen reageren zonder daar eerst uitgebreid over te moeten delibereren, hebben praktische systemen vaak ook een subsysteem dat zuiver 'reactief' is in de betekenis van Rodney Brooks. Deze is faliekant tegen het gebruik van deliberatie in intelligente systemen vanwege de tijd die dit allemaal kost. Er moet ook worden toegegeven dat dit aspect de vroege experimenten met cognitieve robots in de jaren '70 (bijv. 'Shakey' van SRI) een slechte naam heeft gegeven: die robots waren soms niet vooruit te branden en zaten voortdurend 'na te denken' over plannen e.d.. Echter er is inmiddels veel gebeurd met de hardware die immers veel en veel sneller is geworden, en ik meen oprecht dat zoals zo vaak de waarheid in het midden ligt: voor complexe taken is wel degelijk ook deliberatie nodig. Ik geloof daarom in hybride agent-architecturen die zowel een deliberatieve als een reactieve

component hebben! Heel interessant is trouwens het werk van onderzoekers als Aaron Sloman die nog een stapje verder willen gaan, en zich bekommeren over hoe emoties een rol spelen bij het nog intelligenter (of menselijker?) maken van agents.

Multi-agent systemen en agent societies

Een enkele agent op zichzelf is niet zo interessant. Het wordt pas echt nuttig als we diverse agents - met mogelijk diverse vermogens (*capabilities*) - in een systeem kunnen integreren zodat deze alle gezamenlijk aan een gemeenschappelijke taak kunnen werken. Echter, agents zijn per definitie autonoom, d.w.z. ze vertonen een bepaalde graad van autonoom gedrag. Ze bepalen in principe zelf hun doelen en hoe deze te bereiken. Uiteraard, net als in een menselijke gemeenschap (die ook zou kunnen worden opgevat als een natuurlijk multi-agent systeem), moet deze autonomie worden ingeperkt om de gemeenschap leefbaar te maken en te houden, of anders gezegd, om het systeem efficiënt te laten werken aan een taak. Agents moeten elkaar niet onnodig gaan dwarszitten, bijvoorbeeld. Nu zijn er ook wel *agent societies* waar ook competitie een rol speelt, maar zelfs daar geldt dat agents zich toch aan bepaalde normen of wetten moeten houden, anders wordt het systeem onhandelbaar en onvoorspelbaar, en dat zal over het algemeen toch niet de bedoeling zijn van de ontwerper van het systeem. Dus ook hier spelen normen en waarden een rol, net als in onze maatschappij en de politiek, zij het in een misschien meer technische zin dan waar onze minister-president aan denkt! Het gaat erom het autonome gedrag van de agents in het systeem zo te beperken (*constraining*) dat het systeem als geheel effectief zijn taak kan uitvoeren. Een interessante vraag is hoe agents in een agentgemeenschap zich aan de normen of wetten houden. Hier zijn in principe een aantal oplossingen voor, afhankelijk van de autonomie die de agent gegund wordt. De agent kan de normen vertalen in verplichtingen en verboden, waarvoor het zelf de verantwoordelijkheid neemt deze te respecteren op straffe van sancties, of de agent dient zich te houden aan bepaalde protocollen die garanderen dat de normen worden gerespecteerd. In het uiterste geval wordt de omgeving zo gemaakt dat de agent niet anders kan dan zich



volgens de normen gedragen (denk daarbij aan hoe sommige metrosystemen afdwingen dat je niet zonder kaartje meerijsd). Deze oplossingen zijn uiteraard ontleend aan menselijke gemeenschappen, maar kunnen, afhankelijk van de toepassing, ook goed dienst doen bij artificiële agentgemeenschappen. Een belangrijk begrip hierbij is dat van een elektronische institutie, wat een deel van de omgeving van de agent is dat in feite regelt hoe de erin interacterende agents zich aan de normen houden, meestal middels specifieke protocollen.

In het bovenstaande is het begrip 'coördinatie' van groot belang, wat weer gebaseerd is op communicatie tussen agents. 'Agent-communicatie' is een apart subgebied van de agenttechnologie. Ook hier speelt de autonomie van de agent een rol, op verschillende wijzen. Een agent kan bijvoorbeeld niet zomaar worden gecommandeerd (tenzij er een bepaalde machtsverhouding bestaat). In het algemeen zal men agents verzoeken doen die kunnen worden gehonoreerd als dit past in het eigen plan, doel of intentie van de agent. De semantiek van communicatietalen wordt dan ook vaak gegeven in termen van de mentale toestanden van de agents. Verder kan het zijn dat agents hun eigen taal spreken (omdat ze diverse doelen of expertise hebben). Dan spelen zaken als *ontologieën* een rol, wat in de AI betekent: de wijze waarop de wereld om de agents heen geconceptualiseerd wordt. Men kan zich voorstellen dat communicatie een heel lastig (en mogelijk onoplosbaar) probleem is, als deze conceptualiseringen tussen agents heel veel verschillen. Er zijn twee extremen: enerzijds het triviale geval waarin iedere agent dezelfde taal gebruikt en anderzijds het uiterst moeilijke geval waarin iedere agent een totaal verschillende taal gebruikt. Een interessante onderzoeksvraag is om tussen deze extremen interessante intermediaire gevallen te beschouwen waarbij er een zekere maar niet een totale overlap is tussen de talen die de agents gebruiken. Voor deze gevallen kun je dan protocollen verzinnen waarmee de agents toch met elkaar kunnen converseren en elkaar kunnen 'begrijpen'.

Een ander interessant 'sociaal' verschijnsel in multi-agent systemen of agentgemeenschappen waarin met name cognitiewetenschappers zeer zijn geïnteresseerd, is dat van het *emergent gedrag*. Dit is gedrag van agents dat niet als zodanig is gespecificeerd en geprogrammeerd voor de individuele agents, maar dat onder bepaalde omstandigheden optreedt, ontstaand uit de (vele) interacties van (veelal heel veel eenvoudige) agents. Als men multi-agent systemen gebruikt om het gedrag van bijvoorbeeld biologische systemen te simuleren, is dit natuurlijk bijzonder interessant. Voor de *engineer* die een bruikbaar systeem wil bouwen waarvan het gedrag exact gespecificeerd is, is emergent gedrag misschien eerder een hindernis die moet worden vermeden.

Agent-programmeren

Als we het concept agent serieus nemen als iets waarmee we complexe intelligente systemen willen realiseren, dan is het

van het uiterste belang om aan te geven hoe we agents moeten programmeren. De *agent community* is het niet helemaal eens over hoe dit moet. Bij het realiseren van software onderscheidt men diverse fasen. Over het algemeen spreekt men van een analyse-, ontwerp- en implementatiefase. Alhoewel er een zekere consensus is over het gebruik van agentconcepten in de analyse- en ontwerpfasen, is men het niet eens over het gebruik van deze concepten in de implementatiefase. Sommigen zeggen dat een agentgeoriënteerd ontwerp in algemene talen, zoals Java, moeten worden geïmplementeerd. Anderen, en wij in Utrecht zitten in dit kamp, zijn van mening dat ook in de implementatiefase expliciet gebruikt moet worden gemaakt van deze concepten, wat inhoudt dat de te gebruiken programmeertaal features moet bezitten die direct met deze concepten te maken hebben. Dit soort programmeertalen noemt men met recht 'agentgeoriënteerd'.

Agentgeoriënteerde programmeertalen hebben expliciete 'cognitieve' noties zoals *beliefs* en *goals*, zodat men met deze talen direct het agentgeoriënteerde ontwerp van een stuk software kan programmeren. In feite programmeert men de gewenste mentale toestand (of toestandsovergang) van de agent. De eerste, nog wat primitieve agentgeoriënteerde programmeertaal was AGENT0, voorgesteld door de Amerikaanse onderzoeker Yoav Shoham in 1993. In Utrecht hebben we de agentgeoriënteerde taal 3APL geïntroduceerd. In deze taal is het mogelijk om middels een *belief*- en *goalbase* de mentale toestand van een agent te representeren en middels regels is het mogelijk plannen te genereren en te reviseren (op grond van de huidige mentale toestand van de agent). Deze plannen op hun beurt zijn procedurele entiteiten waarvan de executie de mentale toestand (*belief*- en *goalbase*) van de agent kan veranderen. Het gedrag van de agent ontstaat door afwisselend regels toe te passen om plannen te genereren dan wel te reviseren en plannen uit te voeren. Hoe dit precies gebeurt, is afhankelijk van een *interpreter* die in feite de implementatie is van de eerder genoemde deliberatiecyclus van een agent. We zijn op dit moment bezig deze cyclus zelf weer programmeerbaar te maken, zodat de programmeur precies kan bepalen hoe de agent keuzes maakt - op grond van zijn 'kennis' (*belief*) - bij het kiezen van toe te passen planrevisie of generatie-regels en te executeren doelen. Ook willen we er in de toekomst de mogelijkheid van de afhankelijkheid van dit deliberatieproces van emotionele toestanden van de agent incorporeren, zodat het mogelijk wordt zeer flexibele en daarmee hopelijk zeer intelligente agentsystemen te maken.

Toepassingen

Eigenlijk zijn de toepassingen van agentsystemen legio. Waarmee niet wordt gezegd dat alles per se als een agentsysteem zou moeten worden gerealiseerd. Verre van dat. Mijn standpunt is dat alleen taken of problemen die zich goed lenen voor conceptualisering met behulp van agentbegrippen met agenttechnologie zouden moeten worden aangepakt.

Maar daarvan zijn er vele voorbeelden: van intelligente *personal assistants* tot multi-robotsystemen, van patiënten-monitorsystemen in de geneeskunde tot *deep space explorers* in de ruimtevaart, van heterogene kennissystemen tot systemen voor het uitonderhandelen van verkopen, zoals in de *e-commerce*. Er worden agentsystemen ontwikkeld om gedistribueerd om te gaan met informatie en kennis, waarbij de agents hun eigen kennis en taken hebben, en waarbij gecommuniceerd en onderhandeld moet worden met andere agents als ze er alleen niet uitkomen. Toepassingen van dit soort systemen variëren van systemen ter ondersteuning van veilige informatieoverdracht in een 'vijandige' omgeving en luchtvaart-beheerssystemen tot geheel geautomatiseerde markten voor de verhandeling van goederen. Maar ook in de entertainment-industrie wordt in toenemende mate gebruik gemaakt van agenttechnologie: computergames moeten steeds intelligentere figuren bevatten die met agenttechnologie kunnen worden ontworpen, en de gemiddelde bioscoopganger die naar een film als *Lord of the Rings* kijkt, zal zich niet realiseren dat hier bij de animatie van de enorme massa's figuurtjes gebruik gemaakt wordt van een geavanceerde techniek m.b.v. agenttechnologie: al die figuurtjes in de animatie worden in feite geprogrammeerd als autonome agents die, ingedeeld in diverse soorten, verschillend gedrag vertonen, en waarbij bijv. wordt gezorgd dat ze niet met elkaar (virtueel) in botsing komen.

Het deelgebied van de agenttechnologie dat wordt toegepast om intelligente robotsystemen te maken wordt wel cognitieve robotica (*cognitive robotics*) genoemd. Hierbij moet men niet direct denken aan de zeer creatief schilderende robot in een autoproduktielijn zoals in een recent TV-reclamespotje van Citroën te zien was, maar wel aan robots die op zekere wijze autonoom kunnen opereren, zoals fouten diagnosticeren, deze trachten op te lossen en met andere agents communiceren als dit niet lukt. Ook moeten ze zelf plannings kunnen aanpassen bij geconstateerde afwijkingen van de verwachtingen en onderhandelen met andere robots over het gebruik van *resources*, enz. Ik verwacht dat deze ontwikkeling snel verder zal gaan. In het Zweedse Linköping hebben ze een autonoom vliegende helikopterrobot ontwikkeld die taken als het volgen van auto's - zelfs door tunnels - automatisch kan uitvoeren. NASA heeft voor het International Space Station ISS *robot assistants* ontwikkeld die de astronauten kunnen helpen bij het uitvoeren van hun taken aan boord van ISS, Philips denkt aan een *personal robot* voor gebruik in huis en speelgoedfabrikant Berchet ontwikkelt robots die als speelkameraad kunnen dienen voor jonge kinderen. Deze *intelligent companions* zijn een zeer interessante ontwikkeling waarbij allerlei technieken uit de AI (symbolische en subsymbolische, kwantitatieve en kwalitatieve, agenttechnologie, van deliberatie tot emotioneel gedrag, van spraakherkenning en -productie tot perceptie, etc. etc.) moeten worden gecombineerd en geïntegreerd in één systeem. Ik wil dan ook afsluiten met de verwachting dat IA (*Intelligent Agents*) als moderne vorm van AI een mooie toekomst tegemoet gaat!