

KI, verlengde levenscyclus van "knowledge representation and reasoning"

Dr. Richard Starmans, coördinator SIKS

Wat zijn anno 2005 de belangrijke topics/zwaartepunten binnen het AI onderzoek in Nederland? Welke patronen en onderliggende structuur vallen er te onderkennen in de vele disciplines en subdisciplines, richtingen en stromingen die het AI-veld rijk is? Hoe relevant is in dit verband nog de aloude tegenstelling tussen de symbolische en de subsymbolische AI? En hoe relevant de hieraan verwante tegenstelling tussen "logic-based" en "probability-based" benaderingen?

Deze en andere vragen staan centraal in een uitvoerig empirisch onderzoek dat in de periode oktober 2003 - oktober 2005 aan de UU werd verricht onder ruim 200 onderzoekers, werkzaam bij in totaal tien Nederlandse universiteiten en het CWI. Allen verrichtten in deze periode promotieonderzoek op het gebied van informatie- en kennissystemen (IKS) en waren aangesloten bij de landelijke onderzoeksschool voor Informatie- en Kennissystemen (SIKS). Een aantal saillante bevindingen uit dit onderzoek passeert hier kort de revue.

Het verkrijgen van inzicht in aard, structuur en trends binnen het AI-veld is op voorhand geen sinecure. Zo vormt de AI van oudsher een sterk inter- en multidisciplinair werkterrein. Het wordt beoefend aan verschillende faculteiten en kent uiteenlopende onderzoekstradities en referentie-disciplines. Daarnaast toont de groei van het aantal leerstoelgroepen, studenten en promotieprojecten aan dat het AI-wetenschapsgedebied nog steeds aan belang wint. Tot slot laten focus van de grote internationale AI-conferenties en "aims and scope-sections" van gezaghebbende AI-journals zien dat het vak volop in beweging is en voortdurend evolueert. Oude tegenstellingen, uitgangspunten, paradigma's en oriëntaties hoeven anno 2005 dan ook niet langer courant te zijn.

Al met al vormt de AI een complex en dynamisch werkterrein, dat vreemd genoeg relatief zelden aanleiding lijkt te geven tot grootschalig empirisch onderzoek naar aard, structuur en trends.

Uitgaande van het aantal promotieonderzoeken op het gebied van AI dat thans in Nederland plaatsvindt, kan niet anders dan worden geconstateerd dat dit onderzoeksterrein momenteel een bloeiperiode doormaakt. Starmans (UU) onderzocht aard, structuur en trends van het AI-veld aan de hand van ongeveer 200 Phd-projecten die in Nederland de afgelopen twee jaar zijn uitgevoerd. Het aloude onderscheid tussen symbolische en subsymbolische AI blijkt allerminst aan betekenis te hebben ingeboet, maar behoeft wel degelijk enige revisie. Over de opkomst van Computational Intelligence, de verlengde levenscyclus van de klassieke KR en R, de beloften van Agent Technologie en het weven van het Semantisch Web.

Opzet en methode

Bij onze analyse van het AI-veld staan twee uitgangspunten centraal. Allereerst gaan we uit van feitelijk onderzoek, d.w.z. actuele projecten, zoals die recentelijk zijn verricht dan wel plaatsvinden; dus niet van *mission statements*, onderzoeksagenda's, researchprogramma's, strategische beleidsnota's of subsidieprogramma's. Daarnaast kiezen we er nadrukkelijk voor zo veel mogelijk onderzoekers zelf aan het woord te laten, die via een uitvoerige elektronische enquête zelf hun project karakteriseren en positioneren aan de hand van ongeveer 80 "keywords" of "indexterms", ontleend aan conferenties en *journals* uit het IKS-veld, maar ook in hun eigen bewoordingen. Door onderzoekers zelf op een vijfpunts-

schaal te laten aangeven in welke mate zij zichzelf rekenen tot/ associëren met een bepaald onderzoeksveld of -thema, en deze data vervolgens te analyseren, kunnen reikwijdte, betekenis en relevantie van die termen worden vastgesteld; door onder meer factor analyses kunnen eventuele onderliggende structuren tussen items zichtbaar worden. We zien dus af van het coderen, indexeren of classificeren van wetenschappelijke output of inhoudsanalyse door derden. Evenmin willen we uitgaan van door anderen, of van bovenaf opgelegde kaders of classificaties. Indachtig het "meaning is use"-adagium van de *ordinary language philosophy*, krijgen concepten c.q. begrippen hun betekenis in een gebruikscontext en moeten wij dus nagaan hoe de begrippen door de relevante taalgebruikers ook daadwerkelijk (in relatie tot elkaar) gebruikt worden.

Populatie

Aangezien de AI in ons project is ingebed in het IKS-onderzoek in Nederland en de geldigheid en reikwijdte van onderzoeksresultaten mede bepaald worden door de keuze van populatie/steekproefkader, is een korte toelichting hier op zijn plaats. De door de KNAW erkende onderzoeksschool SIKS is een samenwerkingsverband van ruim 350 onderzoekers, alle werkzaam in het IKS-veld in Nederland. Medio jaren 90 werd de school onder het penvoerderschap van de Vrije Universiteit van Amsterdam opgericht door onderzoekers op het gebied van Artificial Intelligence, Databases/Information Systems en Software Engineering. Inmiddels is de school uitgegroeid van 35 promovendi in 1998 tot ruim 160 dit jaar. Deze promovendi

en stafleden zijn voor het overgrote deel werkzaam bij tien deelnemende universiteiten en het CWI en ondergebracht in ruim 35 verschillende leerstoelgroepen. Weliswaar worden met onze aanpak niet alle AI-onderzoekers bereikt, maar het grote aantal respondenten en de landelijke spreiding geven enige waarborg voor een voldoende representatief onderzoek. Een voordeel is ook dat de resultaten van de enquête verder kunnen worden geïnterpreteerd aan de hand van projectgegevens van de aangesloten onderzoekers.

“Keywords in AI”

Om een eerste indruk te geven van de data, tonen we in tabel 1 de meest relevante keywords, d.w.z. al die thema's die door ten minste 25% van de AI-onderzoekers sterk tot zeer sterk relevant worden geacht voor hun onderzoek. De informatie is beperkt, de termen zijn ongelijksoortig en de tabel zegt niets over de onderlinge relaties, structuur en betekenis van de termen, maar laat wel al enige opvallende constatering toe. Zo valt direct op dat Knowledge Representation and Reasoning (KR&R) en Computational Intelligence (CI) met afstand de belangrijkste keywords zijn voor AI-onderzoekers. KR&R wordt doorgaans opgevat als het theoretisch fundament van de “klassieke”, op logica gebaseerde, symbolische AI en neemt van oudsher een sterke plaats in binnen het veld. CI daarentegen, is van recenter datum en wordt doorgaans veeleer geassocieerd met de subsymbolische AI. Ook valt op dat “probability-based” items (*machine learning, intelligent data analysis, Bayesian/probabilistic networks*) hoog scoren in tabel 1 terwijl klassieke “logic-based” items (*commonsense reasoning, non-monotonic logic, model based reasoning, logic programming*) er niet in voorkomen.

Zoals reeds betoogd, zegt een lijst geïsoleerde en ongelijksoortige keywords weinig over de structuur van het veld. Deze wordt wel zichtbaar als we een aantal factoranalyses uitvoeren op voor de AI relevante keywords uit de enquête. Drie aandachtsgebieden tekenen zich hierbij onmiskenbaar af, die we hier als volgt benoemen:

- Computational Intelligence
- Agent Technology and Multi-Agent-Systems
- Metadata and Ontologies for the Semantisch Web

De analyses tonen drie op zich zelfstaande aandachtsgebieden en het overgrote deel van de projecten blijkt moeiteloos in één van deze onder te brengen. Alle aanleiding derhalve de drie aandachtsgebieden hier kort onder de loop te nemen en hen

te profileren mede aan de hand van projectgegevens van de respondenten.

Tabel 1: Relevantie voor AI	%
Knowledge Representation	63%
Computational Intelligence	53%
Machine Learning	43%
Agent systems	39%
Interactive systems	37%
Reasoning under Uncertainty	36%
Intelligent data analysis	35%
Adaptive systems	34%
Knowledge acquisition	31%
Metadata and Ontologies	30%
Agent architectures	29%
Bayesian/probabilistic networks	29%
Statistical simulation & modeling	29%
Reuse of information	27%
Decision support systems	27%
Distributed systems	27%
Cooperative systems	27%
Semantic web	25%

Computational Intelligence

De belangrijkste AI-trend van de laatste jaren is zonder twijfel de sterke opkomst van CI. Ondanks de soms tegenstrijdige definities die onvermijdelijk opduiken in de literatuur, is er toch

betrekkelijke eensgezindheid dat CI primair complementair is aan c.q. een reactie vormt op de “klassieke” AI, waar vooral symbolische kennisrepresentatie (logica, regelgebaseerde expertsystemen) centraal staat, kennis “top-down” wordt gerepresenteerd en dikwijls complexe onderliggende “models

of thought” niet uit de weg worden gegaan. Uitgangspunt is vaak een “high-level” representatie van domeinkennis middels b.v. epistemische en andere modale logica's en geavanceerde inferentiemechanismen. De beperkingen van deze klassieke AI-benadering leidden ertoe dat binnen en buiten de AI onderzoek werd gedaan naar representatie- en analysetechnieken die kennis veeleer “bottom-up” benaderen b.v. door patronen te herkennen en te “leren” uit (vaak) grote datasets (“emergent”!), in plaats van vooraf de structuur van de oplossing te definiëren. CI betreft dan ook met name technieken, die doorgaans subsymbolisch worden genoemd en kennis op een meer “elemen-

“De groei van het aantal leerstoelgroepen, studenten en promotieprojecten toont aan dat het AI-wetenschapsgebied nog steeds aan belang wint.”

“De belangrijkste AI-trend van de laatste jaren is zonder twijfel de sterke opkomst van Computational Intelligence”

tair” niveau representeren, middels b.v. propositionele logica (*decision trees*, *association rules*, etc) of probabilistische modellen. CI-onderzoek is veeleer statistisch/numeriek van aard en legt een sterk accent op computationele aspecten. Een ander opvallende kenmerk is dat bij CI meer de nadruk ligt op *benchmarking*, validatie en toepassing in realistische “real-life” toepassingsdomen.

CI-onderzoekers, scoren hoog op thema’s als Intelligent data analysis, Data mining, Machine learning, Bayesian networks, Neural networks, Pattern recognition, Adaptive behaviour, Statistical modeling, Reasoning under uncertainty en Quantitative empirical research. Ook geven zij aan KR&R zeer relevant te vinden! Toepassingen in onze populatie variëren van bio-informatica, medische beslissingsondersteunende systemen, *information retrieval* en *data mining* voor het semantisch web, tot taal- en spraak technologie en *machine learning* algoritmen om de authenticiteit van schilderijen te bepalen.

Agent technologie en multi-agent systems

Agent technologie en *multi-agent systems* is een tweede aandachtsgebied dat in alle analyses onmiskenbaar naar voren komt. *Agents* zijn hardware of software systemen die autonoom acties uitvoeren in een veranderende omgeving (zie ook het artikel van John-Jules Meyer, De Connectie, jrg1, nummer 1). Te denken valt aan *software agents* die informatie opsporen, verzamelen en aggregeren, op zoek gaan naar specifieke producten op het internet, prijzen vergelijken, een bod doen op een veiling en de transactie tot stand brengen. Veel onderzoek betreft *multi-agent* systemen, waarbij verschillende actoren interactie hebben en wedijveren/onderhandelen dan wel samenwerken als competitieve dan wel coöperatieve *agents*. Een stap verder is als deze *agents* ook sociale groepen of *societies* gaan vormen. *Agent-logics* die kennis en intenties van *agents* representeren, *agent-talen* en -architecturen spelen hierbij een belangrijke rol. Een van de voorlopige hoogtepunten voor het *agent*-onderzoek in Nederland was de AAMAS 2005, die dit jaar in Utrecht werd georganiseerd. Blijkens onze enquête is het *agent*-onderzoek in Nederland sterk geworteld in de klassieke symbolische AI en geeft daarmee een belangrijke nieuwe impuls aan KR&R. Het is overwegend theoretisch van aard en kent in tegenstelling tot CI nog relatief weinig *real-life* toepassingen.

Het weven van het semantisch web

Met behulp van XML en aanverwante technologieën kan webinformatie platformafhankelijk en apparaatafhankelijk worden getoond. Een stap verder gaat het zgn. Semantic Web dat de laatste jaren in de wetenschappelijke literatuur een ware opmars beleefd en met name onder AI-onderzoekers weerklank vindt. De term werd door de “founding father” van het internet, Tim Berners-Lee, zelf geïntroduceerd om een nieuwe generatie internet aan te kondigen, die “machine accessible” is. Door gebruik te maken van metadata en ontologieën wordt

achtergrondkennis toegevoegd aan webpagina’s, die computers in staat stelt informatie te zoeken, te combineren en te interpreteren. Het zoeken naar *strings/keywords* en “screen-scraping” moeten dus plaats maken voor semantiek en “concepten”. Ook technologieën als *document languages* (XHTML, SMIL, SVG, MathML), *document transformation languages* (XSLT, XSL) en Semantic Web *languages* (RDF, RDFS, DAML+OIL, OWL) bieden volop nieuwe uitdagingen. De opmars van het Semantisch Web en de hernieuwde belangstelling voor metadata en ontologieën heeft een nieuwe impuls gegeven aan KR&R. Dat laatste is ook het standpunt van Berners Lee, die meermalen de stand van zaken binnen KR&R heeft vergeleken met de status van *hypertext* vóór de opkomst van het internet; er waren aardige theorieën, kleinschalige implementaties en beloften, maar tot wereldschokkende resultaten kwam het aanvankelijk niet. De “killer-application” diende zich pas aan met de komst van het internet. Volgens Berners Lee kan en moet het Semantisch Web dezelfde rol voor KR&R vervullen, die het internet inderijd voor de *hypertext*-gemeenschap vervulde.

Opvallend is verder dat dit aandachtsgebied niet uitsluitend subsymbolisch is; in een groeiend aantal projecten spelen ook thema’s als *machine learning* en *data mining* een rol.

Conclusie

De sterke toename van de sub-symbolische/*probability-based* AI ten koste van symbolische/*logic-based* AI leidt allerminst tot een teloorgang van KR&R. Enerzijds valt dit te verklaren uit het feit dat KR&R niet uitsluitend betrekking heeft op de klassieke symbolische AI, maar ook door de subsymbolische AI (en zelfs buiten de AI!) zeer relevant wordt bevonden. Paradoxaal genoeg draagt de sterke opmars van CI dus bij aan de dominante positie van KR&R binnen de AI. Anderzijds, leidden de opkomst van de overwegend symbolische “Agent technology and multi-agent-systems” en “Metadata en Ontologies voor het Semantisch Web” wel degelijk tot een verlenging van de “levenscyclus” van de klassieke KR&R. De hiermee gepaard gaande belangstelling voor b.v. logica en klassieke inferentiemechanismen, ontologieën, metadata en *description logics* vormt daarmee een tegenwicht tegen de afnemende aandacht voor b.v. *non-monotonic reasoning*, *commonsense reasoning*, *logic programming*, *modal logics* of *rule-based expertsystems*. Dit alles suggereert ook dat KR&R niet zozeer een apart aandachtsgebied binnen de AI vormt, maar een rol speelt in alle drie in onze data geïdentificeerde aandachtsgebieden. ∅

In deze beknopte bijdrage kan slechts een zeer summier beeld worden gegeven van ontwikkelingen binnen de AI in Nederland. Wie meer informatie wenst kan contact opnemen met de auteur: starmans@cs.uu.nl.