

# Over Machines die Leren

*Luc De Raedt, Departement Computerwetenschappen, Katholieke Universiteit Leuven*

Lessen voor de 21ste eeuw, 24 november 2008

## Inleiding

De kunstmatige intelligentie wil machines (zeg maar, computers of robots) ontwikkelen die intelligent gedrag vertonen. Eén van de centrale aspecten van intelligentie is de mogelijkheid tot leren. Vandaar dat het *machinaal leren* ook één van de belangrijkste deelgebieden van de kunstmatige intelligentie (en de informatica) is. We zeggen dat *een machine leert indien zijn prestaties op een welbepaalde taak beter worden met ervaring*.

Eén van de eerste en nog steeds fascinerende voorbeelden van een lerende machine is het damprogramma van Arthur Samuel, één van de pioniers van de kunstmatige intelligentie. Dit spelprogramma speelde het damspel (de *taak*), en naarmate het meer spelletjes speelde (de *ervaring*) werd het sterker. Het programma werd ontwikkeld in de jaren 50 en 60 en vormde één van de vroegste demonstraties van de kunstmatige intelligentie. Het slaagde er in die tijd al in om te winnen van één van de betere Amerikaanse damspelers.

Machinaal leren is interessant vanuit verschillende perspectieven. Tijdens de les wordt het echter vooral vanuit een praktisch informatica perspectief benaderd. Zo kan men met behulp van machinaal leren systemen bouwen die onmogelijk met de hand kunnen geprogrammeerd worden. Ook laat het machinaal leren toe om op grote schaal gegevens te analyseren om hieruit nieuwe kennis af te leiden. In dit opzicht is het nauw verwant met *data mining*.

## Voorbeelden van Lerende Systemen

Er zijn heel wat voorbeelden van lerende systemen, ook in het dagelijks gebruik, die we kort bespreken. Zo is het filteren van spam bij e-mail gebaseerd op het machinaal leren. De gebruiker van het e-mail programma kan daarbij aangeven welke e-mails ongewenst zijn, en de filter zal die informatie gebruiken om nieuwe e-mails te classificeren als gewenst of ongewenst. Naarmate de gebruiker meer e-mails classificeert, past de filter zich meer en meer aan zijn gebruiker aan en worden de classificaties nauwkeuriger. Het is bijna ondenkbaar om spamfilters te ontwikkelen zonder gebruik te maken van leertechnieken omdat een spamfilter zich moeten aanpassen aan de individuele gebruikers en omdat spam ook constant evolueert. Een tweede voorbeeld gaat over de webstek van internetbedrijven zoals Amazon. De analyse van de aankopen van klanten laat dergelijke bedrijven toe om aanbevelingen en zelfs speciale aanbiedingen te maken voor hun gebruikers. Bv. als klanten interesse hebben voor een bepaald boek, worden meteen suggesties gemaakt voor andere boeken, die vorige klanten samen gekocht hebben. Andere toepassingen vindt men bij de analyse en interpretatie van wetenschappelijke gegevens, waar het vaak makkelijk is om gegevens te genereren, maar moeilijk om ze te interpreteren. Om het met een slogan te zeggen “*We are drowning in data but starving for knowledge*”. Het machinaal leren en het data mining is dan ook prominent aanwezig bij veel wetenschappelijke domeinen zoals de bio- en de chemo-informatica en de computerlinguïstiek.

## Technieken van Machinaal Leren

Het leren uit *beloningen* wordt gebruikt om te leren welke acties systemen best ondernemen in welke situaties. Dit is de techniek die door Arthur Samuel ontwikkeld werd en sindsdien

geperfectioneerd werd. Deze techniek zal geïllustreerd worden a.d.h.v. het MENACE programma (1961) van Donald Michie voor het spelen van Tic-Tac-Toe (boter-kaas-en-eieren). De kernidee is dat acties die een beloning opleveren frequenter uitgevoerd worden. Het probleem bestaat er dan in om te achterhalen welke acties verantwoordelijk zijn voor een beloning, zoals het winnen van een partij.

Het leren van *associaties* wordt gebruikt door Amazon. Uitgaande van de verzameling aankopen van klanten (bv. de gekochte boeken) worden regels berekend van de vorm “*ALS boek X gekocht wordt DAN wordt ook boek Y gekocht*”. Hoe vaker *Y* samen gekocht wordt met *X* hoe relevanter de aanbeveling voor boek *Y* wanneer de klant nu boek *X* bestudeert of aankoopt.

Als derde voorbeeld wordt besproken hoe *beslissingsbomen* kunnen afgeleid worden uit voorbeelden. Dit is nuttig voor *classificatieproblemen*, die bv. optreden bij spamfilters. Een beslissingsboom voorspelt de klasse van een voorbeeld vertrekkende van de eigenschappen van dit voorbeeld (bv. de e-mail wordt als gewenst of ongewenst geclassificeerd door gebruik te maken van de woorden die voorkomen in de e-mail en andere kenmerken zoals de afzender en het onderwerp). Het machinaal leren kan geformaliseerd kan worden als het leren van functies uitgaande van verschillende soorten ervaring (of voorbeelden). Ook onderzoekt men daarbij de algemene wetmatigheden en beperkingen die daarbij geldig zijn.

### **Machinaal Leren Nu en in de Toekomst**

Het machinaal leren wint zowel aan belang als aan erkenning. Er worden meer en meer toepassingen ontwikkeld, die volgens de auteur een grote invloed zullen hebben op zowel het wetenschapsbedrijf als op het dagelijkse leven. Het Robot Scientist project van Ross D. King et al. onderzoekt hoe technieken van machinaal leren en kunstmatige intelligentie gebruikt kunnen worden om het opstellen en het verwerken van wetenschappelijke experimenten te automatiseren. King et al. schetsen een experiment waarbij een lab-robot zelfstandig besliste welke experimenten door te voeren en de resultaten ervan gebruikte om een theorie in een microbiologische toepassing aan te passen. Een illustratie van een totaal andere aard betreft de elektronica die momenteel overal ingebouwd wordt. Die laat toe om een rijk gamma aan informatie over de toestand van de gebruiker en zijn toestel te genereren. Die informatie kan dan op zijn beurt gebruikt worden om de gebruiker te assisteren bij zijn taken. Zo kan verwacht worden dat in de toekomst personen die bv. de verkeerde bus nemen of hun sporttas vergeten op weg naar hun wekelijkse training, hierop gewezen worden door hun elektronische assistent; dat verwarming, licht en andere toestellen in een huis semi-automatisch geregeld en ingesteld worden op basis van de vroegere voorkeuren van hun bewoners; dat ouderen langer alleen zullen kunnen blijven wonen; dat voor routine rapporten van bv. agenten of verplegers er automatisch een voorstel gegenereerd wordt, enz. Deze laatste illustraties behoren tot het domein van *human activity recognition*.

### **Examenvragen**

1. Wat is machinaal leren ? Hoe kan het geformaliseerd worden ? en waarom is het van belang ?
2. Bespreek één techniek van het machinaal leren (leren van beslissingsbomen, uit beloningen, of van associaties). Illustreer je antwoord met een voorbeeld.
3. Bespreek één actuele onderzoekstrend (leren in wetenschappelijke domeinen, het herkennen van menselijke activiteiten, of het onderzoeken van netwerken) binnen het machinaal leren.
4. Geef een eigen voorbeeld van een mogelijke toepassing van machinaal leren. Zeg wat er geleerd wordt, welke ervaring ter beschikking staat, en argumenteer waarom de toepassing interessant is.