

# Fuzzy rules in neurale netwerken en artificiële intelligentie

H.J. Kappen, *Stichting Neurale Netwerken, Laboratorium voor Medische en Biofysica, Nijmegen*

## Inleiding

Artificiële intelligentie staat volop in de belangstelling. Niet omdat het een geheel nieuwe tak van wetenschap is: in de jaren vijftig en zestig was men er al druk mee bezig. De voorspellingen die men toen deed, wekten veel verwachtingen. Daaraan werd maar ten dele voldaan. Sinds de tachtiger jaren echter zijn er ontwikkelingen op gang gekomen en die de verwerking van de dromen uit de jaren zestig een stuk dichterbij hebben gebracht.

Het AI-onderzoek richt zich vooral op eigenschappen als het menselijk vermogen tot redeneren, spreken, handelen, waarnemen en bewegen. Dus zowel hogere functies (kennis en redeneren) als ook de meer algemeen biologische functies, zoals waarnemen en bewegen worden onderzocht.

Op dit laatste gebied boeken neurale netwerken beoedigende resultaten. Het zijn computersystemen die gemodelleerd zijn naar de hersenen. Neurale netwerken kunnen fungeren als een associatief geheugen: dit geheugen stelt het netwerk in staat om op een intelligente manier te redeneren wanneer incomplete of verstoorde informatie wordt aangeboden; ook maakt dit geheugen het leggen van verbanden mogelijk. Daarnaast hebben neurale netwerken het vermogen om te leren. Ze zijn in staat om patronen te leren herkennen en ze bezitten tevens capaciteiten om te abstraheren en te generaliseren. Toepassingen binnen de AI zijn er op het gebied van herkenning van beelden en spraakherkenning, voor het aansturen van de ledematen van robots en voor planning en navigatie van autonome voertuigen. Daarnaast zijn er in de industrie vele patroonherkenningstoepassingen, bijvoorbeeld voor interpretatie en classificatie van multi-variate data, voor het voorspellen van tijdreeksen en voor besturing van industriële processen.

Een belangrijke voorwaarde voor het succesvol toepassen van de (huidige) neurale netwerken is, dat de data die het probleem beschrijven een topologische structuur bezitten. Hiermee wordt bedoeld, dat één kleine verstoring van het probleem tot een kleine verstoring van de oplossing leidt. Aan deze voorwaarde is voor al de bovengenoemde toepassingen voldaan. Daardoor komen alle aantrekkelijke eigenschappen van neurale netwerken, zoals leren, generaliseren, associatief gedrag en robuustheid tegen ruis ter beschikking. Als de data uit signalen bestaan (continue getallen) bezitten ze vrijwel altijd een topologische structuur.

Wat betreft de hogere functies, zoals redeneren en

omgaan met symbolische data, hebben neurale netwerken tot nu toe nog niet veel resultaten geboekt. Dit is voornamelijk te wijten aan het feit dat symbolische data vaak geen topologische structuur bezitten. Bij bijvoorbeeld schaken wordt het neurale netwerk getraind om het verband tussen een bepaalde opstelling op het bord en de optimale volgende zet te leren. Omdat het ondoenlijk (en oninteressant) is om het netwerk op alle mogelijke stellingen te trainen, zal een *representatieve subset* worden aangeboden. Na trainen wordt het netwerk gevraagd om op een onbekende stelling een goede zet te doen. Het netwerk zal zoeken welke hem bekende stelling het meest op deze stelling lijkt en een zet voorstellen die lijkt op de optimale zet van deze bekende stelling. Dit moet mis gaan omdat het bij schaken niet zo is dat bijna gelijke stellingen om bijna gelijke oplossingen vragen.

## Expertsystemen

Typisch symbolische problemen kunnen succesvol worden opgelost met zogenaamde expert systemen. Een expert systeem bestaat uit een aantal regels die kennis over het probleemgebied uitdrukken. Deze regels worden zodanig ontworpen, dat het gedrag van het systeem zo goed mogelijk de menselijke expert benadert. Dergelijke systemen worden ook in de medische wereld toegepast. Er zijn vele voorbeelden te noemen, maar het meest bekende is het stellen van een diagnose aan de hand van een aantal eigenschappen en symptomen van een patiënt. De kracht van dergelijke systemen is dat er vrijwel geen beperkingen zijn aan de uitdrukingskracht van een dergelijk systeem. Alle regels die men kan bedenken kunnen in een expertstelsel worden opgenomen.

Het is echter gebleken, dat het niet eenvoudig is om een dergelijk regelsysteem te maken. Ten eerste blijken menselijke experts moeite te hebben om de regels die zij hanteren expliciet te maken. Zij hanteren algemene regels, die vaak gelden, maar niet altijd. Als de

**Dr. H.J. Kappen** studeerde theoretische natuurkunde aan de Rijksuniversiteit Groningen en promoveerde in 1987 aan de Rockefeller Universiteit in New York. Van 1987 tot 1989 was hij als wetenschappelijk medewerker verbonden aan het Philips Natuurkundig Laboratorium. Sedert 1989 is hij als projectmanager bij de Stichting Neurale Netwerken te Nijmegen werkzaam en actief betrokken bij de research van neurale netwerken.

algemene regel niet geldt, geldt er mogelijk een speciale regel, tenzij... enzovoort. Bijvoorbeeld, het in kaart brengen van alle combinaties van mogelijkheden die tot het ziektebeeld van een patiënt kunnen bijdragen en het vastleggen welke regel in welk geval geldt wordt al snel een uitzonderlijk complexe taak. Ten tweede is het vaak moeilijk harde grenzen aan te geven omtrent het exacte geldigheidsdomein van een regel. Voorbeelden zijn regels als: 'Iemand met koorts kan maar beter binnenshuis blijven'. De dwingendheid van deze regel staat klaarblijkelijk ter discussie. Gedeeltelijk hiermee samenhangend is de definitie van 'koorts'. Is de lichaamstemperatuur hoger dan 38°C of hoger dan 39°C? Waarschijnlijk is de dwingendheid van de regel afhankelijk van de hoogte van de temperatuur. Dit is een voorbeeld van een *fuzzy rule*: er is een fuzzy relatie tussen lichaamstemperatuur en het begrip koorts en er is een fuzzy relatie tussen koorts en het al dan niet binnen blijven. Het probleem met het gebruiken van fuzzy rules voor expertsystemen is dat het de taak van het ontwerpen van het expertstelsel nog veel ingewikkelder maakt, omdat nu ook nog voor elke regel moet worden bepaald in welke mate deze regel geldig is. Ten derde blijkt het vaak, dat de gevallen die overblijven wanneer de belangrijkste regels zijn geïdentificeerd, zich nauwelijks door regels laten beschrijven. In deze geïsoleerde gevallen is het enige wat de menselijke expert kan doen, redeneren door analogie: 'dit ziektebeeld is een uitzonderlijk geval, maar lijkt het meest op dit andere geval, dus we stellen een soortgelijke diagnose'. Redeneren door analogie, is een vorm van associatief denken, die moeilijk in een regelsysteem kan worden ingebouwd. Hier zouden neurale netwerken een bijdrage kunnen leveren.

We zien dus, dat praktische kennisproblemen gekarakteriseerd worden door symbolische aspecten, die redelijk goed door exact geldende regels worden beschreven, en ruis-aspecten die kunnen worden beschreven door fuzzy regels of met associatief redeneren moeten worden aangepakt. Neurale netwerken kunnen aan de bovengenoemde problemen wellicht een belangrijke bijdrage leveren. In zogenaamde Boltzmann Machines (BMs) kunnen zowel associatief gedrag als fuzzy regels worden gerealiseerd. Ik zal mij in deze bijdrage tot dit laatste beperken.

### Neurale Netwerken implementeren fuzzy regels

Beschouw een classificatieprobleem met slechts twee klassen. Denk hierbij bijvoorbeeld aan een neuraal expertstelsel dat aan de hand van patiëntgegevens moet beslissen of de patiënt ziek of gezond is. Laat  $\vec{x}$  een vector zijn die de patiëntgegevens bevat en  $y = \pm 1$  codeert of de patiënt gezond of ziek is. De conditionele kans  $p(y|\vec{x})$  geeft de waarschijnlijkheid aan dat een patiënt met eigenschappen  $\vec{x}$  ziek of gezond is. Het is duidelijk dat elke patiënt ziek of gezond is:

$p(+1|\vec{x}) + p(-1|\vec{x}) = 1$ , voor elke  $\vec{x}$ . De conditionele kans  $p$  implementeert de fuzzy regel: 'als  $\vec{x}$  dan  $y$  met kans  $p(y|\vec{x})$ '.

Het neurale netwerk dat zich het beste voor regelinterpretatie leent, is de zogenaamde Boltzmann Machine (1,2). Het netwerk bestaat uit een aantal neuronen, dat met elkaar verbonden is door middel van synapsen. Langs deze verbindingen ontvangt elk neuron activiteit van andere neuronen. Deze activiteit wordt het lokale veld van een neuron genoemd. Hoe groter het lokale veld, hoe groter de kans dat het neuron vuurt. Een vurend neuron heeft activiteit 1 en een niet vurend neuron heeft activiteit  $-1$ . Deze *stochastische dynamica* wordt gelijktijdig door alle neuronen uitgevoerd. Na verloop van tijd zal zich een evenwicht instellen: als de activiteit van de neuronen gegeven wordt door een vector  $\vec{s}$  dan geldt in evenwicht dat de kans dat een neuron vuurt een constante waarde wordt onafhankelijk van de tijd:

$$p(\vec{s}) = \frac{\exp(-H(\vec{s})/T)}{Z}$$

$Z$  is een normeringsconstante. De Boltzmann verdeling  $p$  is een bekend begrip in de statistische mechanica.  $H$  beschrijft de energie van het systeem. Men ziet dat in evenwicht toestanden met lage energie waarschijnlijker zijn dan toestanden met hoge energie. Tis een constante die hier de mate van ruis in de dynamica van de neuronen regelt: hoe hoger  $T$  des te meer ruis.  $T$  wordt ook wel de temperatuur van het systeem genoemd naar analogie van andere toepassingen. Boltzmann Machines kunnen worden gebruikt voor het leren van conditionele kansen. In het geval dat de conditionele kans  $p(y|\vec{x})$  wordt geleerd bestaat het netwerk uit een laag van inputneuronen waarop de gegevens van de patiënt worden gecodeerd. De tweede laag bevat een aantal ( $h$ ) verborgen neuronen. De output van het netwerk bestaat uit 1 neuron dat de waarde  $y$  codeert (Fig. 1).

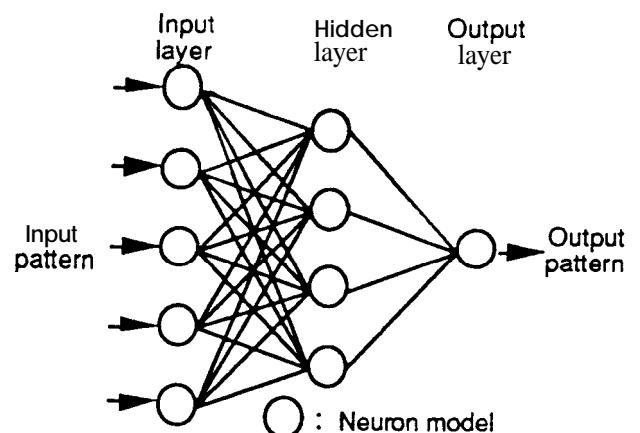


Fig. 1 Neurale netwerk architectuur voor de Boltzmann Machine. Elk verborgen neuron is verbonden met alle inputneuronen en met het outputneuron. Er zijn ook directe verbindingen tussen input- en outputneuronen (niet getekend in de figuur).

De inputneuronen van het netwerk worden vastgezet op een patroon  $\vec{x}$ . De rest van de neuronen relaxeert naar een evenwichtsverdeling. Voor het outputneuron geldt in evenwicht:

$$p(y|\vec{x}) = \frac{\exp(yf(\vec{x})/T)}{\exp(f(\vec{x})/T) + \exp(-f(\vec{x})/T)}$$

De functie  $f(\vec{x})$  hangt af van de gewichten van het netwerk. Door het leerproces worden deze zo gekozen, dat  $p(y|\vec{x})$  zo goed mogelijk de conditionele kansen van de patiënten in de trainingset benadert. De fuzziness van de regel wordt gecontroleerd door  $T$ . Voor  $T = 0$  geldt de deterministische regel 'als  $f(\vec{x}) > 0$  dan  $y = 1$ '. Dus het netwerk voert een drempellogica uit, waarbij de drempel gegeven wordt door de verzameling  $\{\vec{x} | f(\vec{x}) = 0\}$ . Het *fuzzy domein* van de regel omvat die patiënten waarvoor de classificatie onzeker is, bijvoorbeeld:

$$\{|\vec{x}| - 2 < f(\vec{x})/T < 2\}$$

definieert een fuzzy domein, waarbuiten alle patiënten met meer dan 98 % zekerheid worden geclassificeerd.

Men kan een leeralgoritme ontwerpen, dat er voor zorgt dat voor elke patiënt uit de verzameling van trainingspatronen zo veel mogelijk de gewenste output wordt gegeven. De trainingsfout is grofweg gezegd, het aantal misklassificaties op de verzameling van trainingspatronen. Het leeralgoritme probeert de trainingsfout te minimaliseren.

Neem als voorbeeld een neurale netwerk zonder hidden units. Er geldt dan:

$$f(X) = \sum_{i=1}^n w_i x_i - \theta$$

Dit netwerk implementeert de regel

'als  $\sum_{i=1}^n w_i x_i > \theta$  dan  $y = 1$  met kans  $p(y|\vec{x})$ '.

Door het leren wordt de optimale waarden voor  $w_i$  en  $\theta$  gevonden. Merk op dat het leerproces twee dingen leert: de locatie van het beslissingsvlak  $f(\vec{x}) = 0$  en de fuzziness van de classificatieregels. Het eerste wordt gegeven door de *relatieve* grootte van de gewichten en de drempel, het tweede door de absolute schaal van de gewichten en de drempel. Dus de temperatuur  $T$  kan willekeurig gekozen worden, omdat het leerproces automatisch de correcte schaal van het fuzzy domein bepaalt.

Het is duidelijk dat de uitdrukingskracht van dit neurale netwerk slechts afhangt van de mogelijke functies  $f$  waarover het netwerk kan beschikken. Dit blijkt geen beperking te zijn. Men kan bewijzen, dat voor een willekeurige trainingset met  $P$  patiënten  $(\vec{x}_i, y_i), i = 1, \dots, P$  er altijd een  $f$  bestaat die voor elke patiënt correct classificeert, indien men voldoende verborgen neuron  $h$  neemt (3). Anders gezegd, de trainingsfout is een dalende functie van  $h$  en wordt

nul voor een eindige waarde  $h$ .

Het generaliserend vermogen van het netwerk wordt getest, door de uitkomst van het netwerk te vergelijken met de correcte uitkomst op een onafhankelijke verzameling van testdata. Het aantal misklassificaties op de testverzameling heet de testfout. De testfout is voor kleine  $h$  ook een dalende functie van  $h$ . Echter, voor grote  $h$  zal de testfout toenemen met  $h$ . Dit effect wordt veroorzaakt doordat het netwerk zich heeft (over)gespecialiseerd op de trainingset en hierover details heeft geleerd die specifiek zijn voor die trainingset en geen algemene trend in het probleem vertegenwoordigen.

### Ovariumcarcinoom

In samenwerking met het Academisch Ziekenhuis van de Rijksuniversiteit Universiteit Utrecht wordt in Nijmegen onderzoek gedaan naar het voorspellen van de overlevingskansen en de responsie op therapie van patiënten met een ovariumcarcinoom. In dit onderzoek worden neurale netwerken gebruikt voor de analyse van de effectiviteit van verschillende vormen van chemo-therapie. Daarnaast zijn we geïnteresseerd in het identificeren van de belangrijkste voorspellende factoren. Tevens wordt geprobeerd om uit de structuur van het neurale netwerk regels af te leiden.

De data die gebruikt zijn komen van twee klinische studies van de Nederlandse Studiegroep voor Ovarian Cancer. De eerste studie begon in 1979 en vergeleek twee kuren genaamd Hexa-CAF en CHAP-5. De tweede studie begon in 1981 en vergeleek CHAP-5 met CP. De evaluatiedatum voor beide studies was 1 januari 1989. Het totaal aantal patiënten in beide studies was 377. Per patiënt zijn een aantal kenmerken bekend, zoals leeftijd, gewicht, uitgebreidheid en ernst

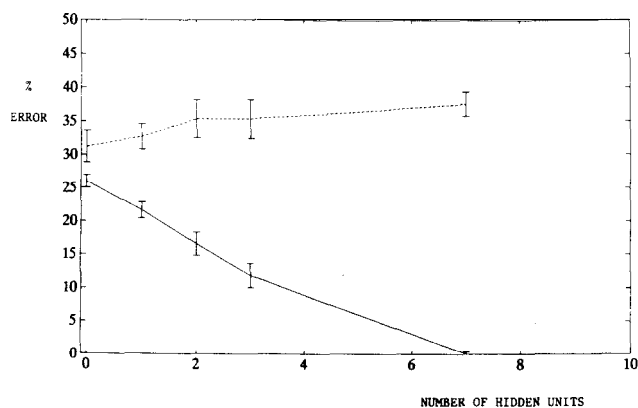


Fig. 2 Trainingsfout (onderste lijn) en testfout (bovenste lijn) als functie van het aantal hidden units. De taak van het netwerk was om te voorspellen of patiënten langer of korter dan 2 jaar overleven gebaseerd op patiënten- en behandelingsgegevens.

van de tumor en fysieke conditie. Daarnaast zijn er een aantal kenmerken die de behandeling van de patiënt beschrijven, zoals de soort therapie, de medicijnen die zijn toegediend en het aantal therapieën.

Wij hebben een Boltzmann Machine getraind om op basis van patiënten- en behandelingsgegevens te voorspellen of patiënten langer of korter dan 2 jaar zullen leven. De grens van 2 jaar was gekozen opdat beide klassen circa 50 % van de patiënten bevatten. Hierbij hebben we het aantal hidden units gevarieerd. 75 % van de patiënten werd steeds gebruikt om het netwerk te trainen en 25 % om het generaliserend vermogen te testen. Dit werd 20 keer herhaald, om over de specifieke keuze van training- en testset te randomiseren. De resultaten zijn weergegeven in figuur 2.

Uit figuur 2 blijkt, dat ondanks de hoge trainingsfout het optimale aantal hidden units voor dit probleem bij 0 ligt. Men kan hieruit de conclusietrekken, dat de combinatie van de complexiteit van het probleem en de geringe hoeveelheid beschikbare data een regelextractie met een complexer neurale netwerk niet toestaat.

We kunnen een soortgelijke analyse ook doen alleen op basis van gegevens die bekend zijn voor de start van de behandeling. Ook dan blijkt een netwerk zonder verborgen neuronen de beste generalisatie te geven (circa 68 %).

We kunnen de regel die het netwerk geleerd heeft nu interpreteren. We definiëren voor elke patiënt de prognostische index zoals hierboven:

$$f(\vec{x}) = \sum_{i=1}^n w_i x_i - \theta$$

De kans dat een patiënt langer dan 2 jaar leeft wordt gegeven door  $p(y|\vec{x})$ . Een interpretatie van deze prognostische index laat onder andere zien, dat de conditie en het gewicht van de patiënt een positieve invloed hebben op de levensverwachting en tumor-grootte, uitgebreidheid van de tumor en leeftijd een negatieve invloed hebben. Daarnaast spelen een aantal behandelingskenmerken een belangrijke rol.

### Conclusies

Neurale netwerken kunnen een belangrijke bijdrage leveren voor het ontwerpen van fuzzy expert systemen. Ten eerste, leert het neurale netwerken de relatie

tussen patiënten gegevens en de te stellen diagnose door het presenteren van een aantal voorbeelden. Dit kan, in het geval dat regels niet of slecht bekend zijn, de kwaliteit van het systeem sterk vergroten. De regels die het neurale netwerk gebruikt zijn fuzzy: de voorspelling van het netwerk is een kans om tot een bepaalde klasse te behoren. Tijdens het leerproces worden zowel de beslissingsgrenzen van de verschillende klassen geleerd, als de *fuzziness* van de betreffende regel.

De structuur van het netwerk kan, na het leren worden geïnterpreteerd als een (verzameling) fuzzy regels. In het geval dat het netwerk verborgen neuronen bevat is de interpretatie aanzienlijk moeilijker dan zonder verborgen neuronen.

Ook is het mogelijk, om een bekende verzameling regels in een neurale netwerk te implementeren. Het resulterende netwerk kan vervolgens met een aantal voorbeelden worden doorgeleerd. Indien deze verzameling voorbeelden een representatief aantal typische gevallen (waarvoor regels gelden) en uitzonderingen bevat, zullen de geïmplementeerde regels door het leerproces evolueren tot de gewenste fuzzy regels.

We hebben deze methode geïllustreerd op een medisch probleem. De gepresenteerde resultaten zijn voorlopig en de analyse zal nog moeten worden verbeterd. Op deze wijze kan in het inzicht verkregen worden in de statistische significantie van de gewichten van het netwerk na leren.

### Referenties

- 1 Ackley D, Hinton G and Sejnowski T (1985) A learning algorithm for Boltzmann machines, *Cognitive Science* 9: 147-169
- 2 Kappen HJ (1993) Using Boltzmann Machines as Perceptrons, submitted to *Neural Computation*
- 3 Nilsson N (1965) *Learning Machines*, McGraw-Hill New York

Correspondentie:

Dr. H.J. Kappen

Stichting Neurale Netwerken

Laboratorium voor Medische en Biofysica

Geert Grooteplein Noord 21

6525 EZ Nijmegen