

Kansrijke methoden voor kunstmatig leren en redeneren

H.J. Kappen en M.A.R. Leisink

Stichting Neurale Netwerken,
Laboratorium voor Biofysica
Katholieke Universiteit Nijmegen

Intelligentie is een eigenschap die wij onszelf graag toeschrijven en waarmee wij ons onderscheiden van machines. Kenmerkende eigenschappen van intelligent gedrag zijn het vermogen om te redeneren op basis van onvolledige of tegenstrijdige informatie en het leren uit voorbeelden. Als we in staat zijn deze vermogens na te bootsen, dan kunnen we 'intelligente' machines bouwen. In dit artikel wordt beargumenteerd dat de klassieke kansrekening de meest voor de hand liggende manier is om eenvoudige vormen van leren en redeneren te representeren. Een nadeel is echter dat de tijd die nodig is om de berekeningen uit te voeren exponentieel toeneemt met de probleemgrootte. Mean field theorie, zoals ontwikkeld binnen de statistische mechanica, biedt de mogelijkheid om efficiënte benaderingsmethoden te ontwikkelen en is van essentieel belang om redeneren en leren op basis van kansrekening in de praktijk haalbaar te maken.

1 Probabilistisch redeneren

Hoewel het idee 'kunstmatige intelligentie' al heel oud is, is het met name tot ontwikkeling gekomen na de tweede wereldoorlog. In de jaren '50 werden de eerste schaakcomputers ontwikkeld en ook werden voor het eerst wiskundige stellingen met de computer bewezen. Op basis van deze vroege successen werd door menigeen gedacht dat andere intelligente taken, zoals bijvoorbeeld spraakherkenning en bewegingssturing, relatief eenvoudig met computers zouden kunnen worden opgelost.

De werkelijkheid pakte echter anders uit. De aanpak die de vroege kunstmatige intelligentie programma's volgden, bleek niet bruikbaar te zijn voor grotere problemen. Het aantal mogelijke stellingen dat een schaakprogramma moet analyseren alvorens een optimale zet te doen neemt exponentieel toe met het aantal zetten dat men vooruit kijkt. Dit probleem is kenmerkend voor zeer vele combinatorische problemen en werd herondekt in zeer diverse onderzoekgebieden.

De gemeenschap die zich bezig hield met kunstmatige intelligentie zag zich in het midden van de jaren '60 geconfronteerd met de vraag: Als computers blijkbaar exponentieel veel tijd nodig hebben om tot 'intelligente' oplossingen te komen, hoe doen mensen dit dan? Het antwoord hierop werd gezocht met het idee dat mensen beschikken over 'kennis' die ze gebruiken om efficiënt te zoeken.

In plaats van alle mogelijkheden na te gaan, worden alleen die alternatieven bekeken die relevant lijken. De consequentie was duidelijk. Vraag aan experts welke regels ze gebruiken om problemen op te lossen en combineer deze regels eventueel met andere zoekstrategieën om tot een goede oplossing te komen. Deze aanpak is redelijk succesvol gebleken. Twee bekende voorbeelden van succesvolle expertsystemen zijn MYCIN [Shortliffe and Buchanan, 1975] voor diagnose van bacteriële infectieziekten en PROSPECTOR [Duda et al., 1976] voor winning van mineralen.

Ervaring met het ontwerpen van expertsystemen leerde dat dit geen eenvoudige taak is. Een groot probleem is dat regels niet altijd waar of onwaar zijn, maar een zekere mate van onzekerheid bevatten, en, wat nog erger is, afhangen van de 'context' waarin ze worden gebruikt (oftewel van de waarheid van allerlei andere variabelen). Bijvoorbeeld, stel dat je ziet dat het grasveld in je tuin nat is. Je zal wellicht concluderen dat het geregend heeft. Echter, deze conclusie zal onwaarschijnlijker worden wanneer je waarneemt dat je buurman zijn auto staat te wassen. Dus de relatie of regel "nat gras" \leftrightarrow "regen" hangt af van de context "buurman wast auto".

Dit roept de vraag op wat het correcte formalisme is om met dergelijke onzekerheid om te gaan? Met andere woorden, als we de onzekerheid omtrent de waarheid van een propositie A kwantificeren met een getal $p(A)$, hoe kunnen we dan nieuwe onzekerheden uitrekenen? Dit is een vraag die zijn eigen lange geschiedenis heeft [Jaynes, 1983]. Laplace formaliseerde omstreeks 1774 de kansrekening op basis van ideeën van Bernouilli en Bayes (zie kader 'regel van Bayes').

Regel van Bayes

Als A , B en C gebeurtenissen of uitkomsten voorstellen, dan geldt:

$$p(AB|C) = p(A|BC)p(B|C) \quad (1)$$

$$p(A|B) + p(\bar{A}|B) = 1 \quad (2)$$

Hier beschrijft $p(A|B)$ een kansverdeling over de variabele A voor een gegeven waarde van B . Dus $\sum_A p(A|B) = 1$ voor elke waarde van B .

Omdat AB hetzelfde is als BA kunnen we (1) ook schrijven met A en B verwisseld, zodat:

$$p(A|BC) = \frac{p(B|AC)p(A|C)}{p(B|C)} \quad (3)$$

Deze laatste uitdrukking staat bekend als de regel van Bayes. Hierin verwijst A bijvoorbeeld naar een uitspraak binnen een fysische theorie (kortweg: model), B is de data die we uit een meting verkrijgen, en C bevat alle informatie die we over A hebben voordat we de data B kregen. De regel van Bayes zegt dat de kans op het model gegeven de data niet alleen afhangt van de kans hoe goed het model de data voorspelt, $p(B|AC)$, maar ook van de kans op het model *voordat* we de data zagen, $p(A|C)$.

Dat de regel van Bayes (1-3) de enige juiste manier is om met onzekerheid om te gaan was lange tijd omstreden. In 1946 liet R.T. Cox echter zien dat als we onzekerheid representeren met een reëel getal en als we eisen dat de calculus associatief is (dat als er twee volgorden van berekening bestaan, ze hetzelfde resultaat geven), de resulterende calculus equivalent is met (1-3) [Cox, 1961] (zie ook [Schrödinger, 1947, Reichenbach, 1949]). Dit betekent dat de enige correcte manier om te redeneren met onzekerheid gebaseerd moet zijn op de regels van de kansrekening.

2 Leren en redeneren

Redeneren met kansmodellen betekent het uitrekenen van de kans op een deelverzameling van variabelen A , gegeven dat een andere deelverzameling van variabelen de waarde B heeft. Dat de benodigde rekentijd daarbij exponentieel toeneemt komt doordat over alle overige variabelen moet worden gesommeerd (of geïntegreerd als het continue variabelen betreft). We noemen een algoritme snel (langzaam) als de benodigde rekentijd polynomiaal (exponentieel) toeneemt met de grootte van het probleem. Redeneren in kansmodellen is dus langzaam en praktisch alleen mogelijk voor kleine modellen.

In het algemeen wordt het model $p(\vec{x})$ geparametriseerd door een aantal parameters \vec{w} : $p(\vec{x}|\vec{w})$. (Bijvoorbeeld voor een Gaussisch model specificceert \vec{w} het

gemiddelde en de variantie). In het geval van *leren* worden de waarden van deze parameters bepaald met behulp van een verzameling trainingsvoorbeelden. Elk trainingsvoorbeeld is een patroon \vec{x} en het leeralgoritme past de parameters \vec{w} zo aan dat de kans $p(\vec{x}|\vec{w})$ zo groot mogelijk is, gemiddeld over de gehele verzameling trainingsvoorbeelden. Dit gebeurt doorgaans door het maximaliseren van de zogenaamde likelihood. Dit is de kans op het voorkomen van de trainingsvoorbeelden gegeven het model : $L(\vec{w}) = \prod_{\mu} p(\vec{x}^{\mu}|\vec{w})$ met betrekking tot \vec{w} . μ labelt de trainingspatronen.

In de meeste gevallen zal een leervoorbeeld $\vec{x} = (x_1, \dots, x_n)$ niet elke variabele x_i specificeren (missende waarden). Denk bijvoorbeeld aan een model voor medische diagnostiek. De patientengegevens specificeren voor elke patient de bekende variabelen. Dus $\vec{x} = (\vec{x}_v, \vec{x}_h)$ waarbij \vec{x}_v de bekende variabelen beschrijft en \vec{x}_h de rest. Deze deelverzameling van variabelen is verschillend van patient tot patient. Leren betekent in dit geval dat de parameters \vec{w} zodanig worden gekozen dat de kans op de bekende data $p(\vec{x}_v|\vec{w})$ gemaximaliseerd wordt. Om $p(\vec{x}_v|\vec{w})$ te berekenen dienen alle variabelen in \vec{x}_h te worden uitgesommeerd. Dit is een reden dat leeralgoritmen voor kansmodellen langzaam kunnen zijn.¹

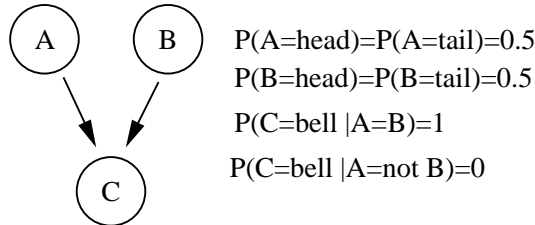
¹Leeralgorithmes kunnen ook langzaam zijn omdat de berekening van $p(\vec{x}|\vec{w})$ zelf langzaam is. Dit is bijvoorbeeld het geval bij Boltzmann verdelingen.

Bayesiaanse netwerken

In Bayesiaanse netwerken wordt het model beschreven als een product van voorwaardelijke kansen:

$$p(\vec{x}) = \prod_i p(x_i | \text{ouders}(i))$$

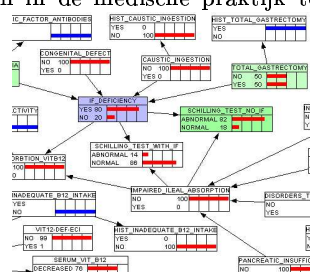
waarbij $\text{ouders}(i)$ een deelverzameling van de variabelen x_1, \dots, x_{i-1} is. Elke kansverdeling kan in deze vorm worden geschreven. De structuur kan grafisch worden weergegeven door een netwerk met variabelen als knopen en conditionele afhankelijkheden als pijlen. Deze aanpak is met name geschikt als de structuur van het netwerk (dwz. welke pijlen aanwezig zijn en welke richting ze hebben) gespecificeerd dient te worden op basis van expertkennis. De reden is dat experts gewend zijn om in termen van oorzaak en gevolg te denken. Deze kennis laat zich goed met voorwaardelijke kansen modelleren. Een voorbeeld is het netwerk in onderstaande figuur.



Twee munten A en B worden onafhankelijk gegooid. Wanneer de uitkomst gelijk is, klinkt er een bel, anders niet. In dit eenvoudige voorbeeld kan men al een gevoel krijgen voor de complexe afhankelijkheden die in dergelijke kansmodellen kunnen ontstaan. Hoewel A en B onafhankelijk zijn, worden ze plotseling volledig gecorreleerd wanneer C bekend is. Dus als we C het gevolg noemen en A en B de oorzaken zien we dat het waarnemen van het gevolg er toe leidt dat er complexe afhankelijkheden tussen oorzaken kunnen ontstaan. Dit is de reden dat redeneren in kansmodellen langzaam is.

Medische diagnostiek

In samenwerking met het academisch ziekenhuis in Utrecht wordt een medisch diagnostisch systeem gebouwd voor (een deel van) het medische domein interne geneeskunde. Hierbij wordt gebruik gemaakt van Bayesiaanse netwerken waarmee de kennis van medische experts wordt gemodelleerd. Het model bevat momenteel ongeveer 80 variabelen en zal naar verwachting in een paar jaar uitgroeien tot ongeveer 1000 variabelen. Het netwerk is nu nog klein genoeg om met exacte methoden te kunnen rekenen. Dat zal niet meer mogelijk zijn wanneer het netwerk een paar honderd variabelen heeft. Het is een belangrijke onderzoeksvraag om na te gaan of mean field methodes voldoende nauwkeurig zijn om in de medische praktijk te kunnen worden toegepast.



Grafisch model voor medische diagnostiek. In de figuur is slechts een klein stuk van het model zichtbaar. De blokken representeren variabelen en pijlen representeren conditionele afhankelijkheden. De rode histogrammen geven de (gemarginaliseerde) kansen van variabelen aan. De blauwe histogrammen geven de waarde van extern ingevoerde variabelen aan. In groen wordt de mate van correlatie van variabelen met de (geselecteerde) lichtblauwe knoop aangegeven.

Een demonstratieversie van dit systeem is beschikbaar op CDrom. Voor meer informatie zie <http://www.mbfys.kun.nl/snn>.

3 Mean field theorie

Het is mogelijk om efficiënte benaderingsmethoden te ontwikkelen voor leren en redeneren. Er zijn ruwweg twee mogelijke manieren van aanpak. De eerste is door gebruik te maken van Monte Carlo samplingtechnieken. Hierbij worden de exponentieel grote sommaties benaderd door alleen de grootste bijdragen op te tellen. Dit kan omdat alle termen in de som positief zijn. De grootste termen worden gevonden door 'op een slimme manier' uit de kansverdeling te trekken. De tweede methode is gebaseerd op mean field theorie. In principe zijn mean field methodes toepasbaar op elk soort kansverdeling. Als voorbeeld beschouwen we de Boltzmann-Gibbs verdeling (zie kader). De Boltzmann-Gibbs verdeling wordt veel gebruikt voor het modelleren van allerlei thermodynamische systemen. In de neurale netwerk literatuur wordt deze verdeling gebruikt voor het beschrijven van het evenwichtsgedrag van een netwerk van neuronen.

Boltzmann verdeling

Naast de Bayesiaanse netwerken bestaan er vele andere kansmodellen die voor redeneren en leren kunnen worden gebruikt. Een interessante klasse van modellen zijn de zogenaamde Boltzmann-Gibbs verdelingen:

$$p(\vec{x}) = \frac{1}{Z} \exp\left(\frac{1}{2} \sum_{ij} w_{ij} x_i x_j + \sum_i \theta_i x_i\right). \quad (4)$$

Hier zijn x_i binaire variabelen ($x_i \in \{0, 1\}$), en w_{ij} en θ_i zijn parameters in het model. Z is de zogenaamde partitiefunctie en zorgt er voor dat de kansverdeling genormeerd is: $Z = \sum_x \exp(\frac{1}{2} \sum_{ij} w_{ij} x_i x_j + \sum_i \theta_i x_i)$. In de statistische fysica beschrijft x_i de toestand van een binaire spin. De Boltzmann verdeling is de verdeling die de entropie van het systeem maximaliseert bij constante gemiddelde energie. In de neurale netwerk literatuur beschrijft x_i het al dan niet vuren van een neuron. Neuronen vuren stochastisch en hun kans op vuren is afhankelijk van de activiteit van de andere neuronen in het netwerk: $p(x_i = 1) = \frac{1}{2}(1 + \tanh(\sum_{j \neq i} w_{ij} x_j + \theta_i))$. De evenwichtsverdeling behorend bij deze stochastische dynamica is de Boltzmann verdeling indien de verbindingen tussen neuronen symmetrisch is: $w_{ij} = w_{ji}$. Deze neurale netwerken staan bekend als Boltzmann Machines.

In dit artikel zullen we ons tot de behandeling van de mean field theorie voor Boltzmann verdelingen beperken [Kappen and Rodríguez, 1998]. De variabelen zullen we ook wel als neuronen aanduiden. Men vervangt het stochastisch gedrag van een neuron (dat gecorreleerd is met alle andere neuronen) door zijn gemiddelde gedrag. Om dat gemiddelde gedrag uit te kunnen rekenen, neemt men aan dat het neuron gekoppeld is met al zijn burens, maar dat deze burens zelf niet meer gekoppeld zijn. Hoewel deze aanname strikt genomen onjuist is, blijkt zij vaak toch tot redelijke resultaten te leiden. De vergelijkingen die men op moet lossen om deze gemiddelde vuuractiviteit van een neuron te vinden, staan bekend als de mean field vergelijkingen. Omdat ‘wel vuren’ wordt gecodeerd door een 1 en ‘niet vuren’ door een 0, geven de mean field vergelijkingen voor elk neuron een getal tussen de 0 en 1 (de gemiddelde vuuractiviteit).

Een directe uitbreiding van de mean field theorie leidt tot de TAP-benadering (genoemd naar Thouless, Anderson en Palmer). De mean field vergelijkingen krijgen dan een correctieterm, die bekend staat als de Onsager reactieterm. Met name wanneer er zowel positieve als negatieve verbindingen tussen neuronen voorkomen, is de oplossing van de TAP-vergelijkingen beter dan de mean field oplossing. Als de koppelingen tussen de neuronen niet al te groot zijn, is de oplossing in de limiet van een oneindig netwerk zelfs exact.

Het is mogelijk deze benaderingen te gebruiken voor het leren van en het redeneren met kansmodellen. Hoewel deze modellen groot kunnen zijn vanuit het oogpunt van de gebruiker (100 - 10000 neuronen), zijn ze nog altijd klein

vanuit een thermodynamisch perspectief. Deze eindige- n -effecten manifesteren zich als correlaties tussen variabelen. Daarnaast is van de homogeniteit in verbindingsstructuur en sterkte, die bij de meeste fysische systemen geldt, bij toepassingen in de AI geen sprake. Dus wanneer we de TAP-vergelijkingen toe willen passen op redeneren in kansmodellen zullen ze hooguit een benadering geven.

In praktische toepassingen is het natuurlijk belangrijk te weten wanneer de gemaakte benadering wel of niet goed is. Omdat de exacte oplossing niet bekend is, is het niet mogelijk direct de benaderingsfout te bepalen. Gelukkig is er een ander criterium, dat in zekere mate bepaalt wanneer de fout ‘groot’ is en wanneer ‘klein’. We zullen dit eerst illustreren met een eenvoudig voorbeeld.

Het is bekend dat

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \dots \quad (5)$$

Dit volgt bijvoorbeeld direct uit de Taylor-ontwikkeling van de functie aan de linkerkant. Deze benadering is echter alleen goed zolang $|x|$ kleiner is dan één. Het is eenvoudig in te zien dat voor $x = 2$ het verschil tussen de linker- en rechterkant alleen maar toeneemt naar mate we rechts meer termen meenemen. Omdat $x = 1$ precies het omslagpunt is tussen ‘wel convergeren’ en ‘niet convergeren’ van de rechterkant, heeft deze benadering een convergentiestraal van één. Deze convergentiestraal valt ook precies samen (en dat is altijd het geval) met het singuliere punt van de linkerkant van de vergelijking.

Ook de mean field en de TAP-benadering zijn te schrijven als een Taylor-benadering in de gewichten w_{ij} van de partitiefunctie. Mean field is hierbij een eerste orde benadering en TAP een tweede orde. Net als in het voorbeeld nemen we als criterium dat de benadering ‘goed’ is als we binnen de convergentiestraal blijven (dus: kleine gewichten). Het is van belang in te zien dat we de fout alleen maar kwalificeren als ‘klein’ of ‘groot’, zonder deze begrippen nader in te vullen. Omdat we de exacte fout niet kunnen berekenen, is dit tot nu toe het maximaal te behalen resultaat.

In figuur 1 is te zien hoe deze methode in de praktijk toch redelijk aangeeft waar de fouten de pan uit beginnen te rijzen. Voor een klein netwerk van 14 neuronen zijn de gewichten langzaam vergroot, zoals op de x -as staat aangegeven. Op de y -as staat fout in de benadering van de correlaties tussen neuronen, die voor dit kleine netwerk nog exact te berekenen is. De verticale stippellijn komt overeen met de convergentiestraal. We zien inderdaad dat buiten deze convergentiestraal de fout ineens fors groter wordt.

4 Het shifter-probleem

Een praktisch probleem dat kan worden geleerd met een Boltzmann machine, is het zogenaamde shifter-probleem. Een patroon van acht bits wordt één positie

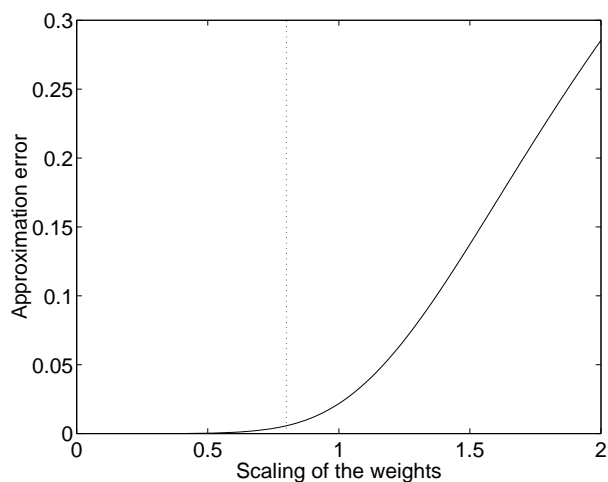


Figure 1: Kwadratische fout in de correlaties volgens de TAP benadering als functie van de grootte van de gewichten ρ .

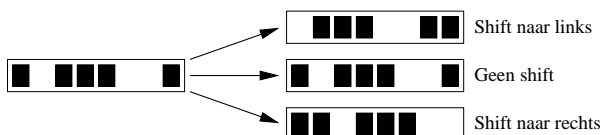


Figure 2: Het shifter probleem. Het netwerk moet herkennen met welke shift (links, geen, rechts) twee patronen aan elkaar gelijk worden.

naar links, naar rechts of niet verschoven, zoals in figuur 2. De bit die eruit wordt geschoven, verschijnt weer aan de andere kant.

Hoewel Boltzmann machines in principe hardwarematig gebouwd zouden kunnen worden, simuleren we het gedrag van de Boltzmann machine met een computerprogramma. De neuronen zijn hierbij eenvoudigweg variabelen in een wiskundig algoritme, die de waarde nul of één kunnen hebben. De gemiddelde vuuractiviteit van een neuron (een tijdsgemiddelde) ligt daarom tussen de nul en één. We gebruiken de eerder genoemde mean field benadering om dit gemiddelde vuurgedrag van de neuronen te bepalen. Eerst leren we de Boltzmann machine door alle 768 patronen één voor één te laten zien. Acht neuronen krijgen het basispatroon te zien; acht neuronen het al dan niet verschoven patroon. Ten slotte krijgt een groepje van drie neuronen informatie over het soort verschuiving. Volgens een bepaald leeralgoritme past de computer de gewichten en drempels van het netwerk aan, zodat het straks in staat is de getoonde patronen zo goed mogelijk te reproduceren. Het gebruikte netwerk is een derde orde Boltzmann machine (zie figuur 3). Het heeft naast de getekende paars-

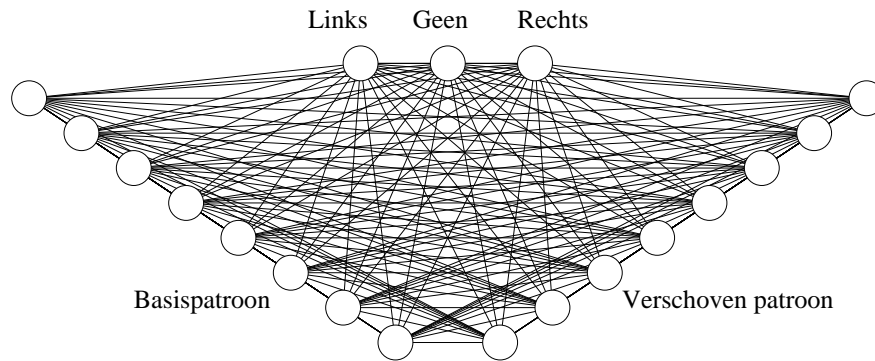


Figure 3: De Boltzmann Machine voor het shifter probleem.

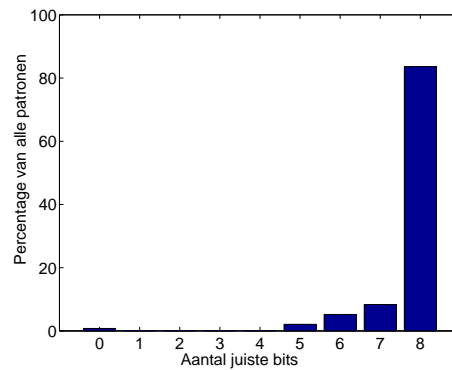


Figure 4: Resultaat op de trainingsset van de Boltzmann Machine voor taak 2.

gewijze verbindingen ook nog verbindingen tussen elk drietal neuronen, omdat de derde-orde-correlaties in dit probleem erg belangrijk zijn.

Nadat het netwerk getraind is, gebruik makend van de mean field benadering, krijgt het twee taken:

1. Het basispatroon en het al dan niet verschoven patroon worden aangeboden; het netwerk moet bepalen welke verschuiving heeft plaatsgevonden.
2. Het basispatroon en de verschuiving worden aangeboden; het netwerk moet zelf het verschoven patroon genereren.

De input-neuronen worden op hun waarde gezet overeenkomstig het gekozen patroon. Het gemiddelde vuurgedrag van de output-neuronen wordt berekend met mean field theorie. De eerste taak werd voortreffelijk uitgevoerd. Alle



Figure 5: Voorbeeld van het gedrag van de Boltzmann Machine op taak 2. a) Een basispatroon; Vereiste shift is naar rechts b) Het voorspelde verschoven patroon

verschuivingen werden correct herkend, behalve de niet-determineerbare verschuivingen (bijvoorbeeld als alle bits in het basispatroon en het verschoven patroon ‘aan’ zijn). De tweede taak, het genereren van een verschuiving, is moeilijker voor het netwerk, omdat het meer informatie zelf moet invullen (acht neuronen in plaats van drie). In het staafdiagram in figuur 4 is af te lezen hoeveel bits van het verschoven patroon correct werden weergegeven. In iets meer dan 80% van de gevallen kon het netwerk het juiste patroon genereren. In figuur 5 staat een voorbeeld van de uitvoer van het netwerk. Links het gekozen basispatroon, dat door het netwerk naar rechts verschoven moest worden. Rechts de gemiddelde vuuractiviteit van de output-neuronen. Als we stellen dat een gemiddelde vuuractiviteit groter dan 0.5 correspondeert met 1 en kleiner dan 0.5 met 0, dan werden in dit geval alle bits op hun juiste waarde gezet door het netwerk.

Dit voorbeeld is nog net klein genoeg om ook het exacte leeralgoritme, dat niet gebruik maakt van de mean field benadering, te kunnen toepassen. Het blijkt dat de bovengenoemde resultaten niet noemenswaardig afhangen van het gebruikte algoritme. We kunnen dus concluderen dat de mean field benadering bruikbaar is voor dergelijke leerproblemen.

5 Discussie

In dit artikel hebben wij gezien dat de klassieke kansrekening de aangewezen methode is voor het modelleren van kennissystemen. Een belangrijk verschil met bijvoorbeeld een (fuzzy) regelsysteem is dat kennis is opgeslagen in conditionele kanstabellen in plaats van in regels. Dit lijkt wellicht op het eerste gezicht een detail, maar zoals we hebben gezien heeft het verstrekking gevolgen. In kansmodellen hebben alle variabelen direct of indirect invloed op elkaar terwijl dit in regelsystemen veel minder het geval is. Dit suggereert dat regelsystemen efficiënter zijn dan kansmodellen. Dit is echter niet geheel correct. Regelsystemen zijn equivalent aan kansmodellen wanneer een aantal aannamen over conditionele onafhankelijkheid worden gemaakt. Dit zijn dus impliciete aannamen waarvan de juistheid sterk afhangt van het toepassingsgebied. Het is in onze optiek daarom beter om dergelijke aannamen expliciet te maken binnen een probabilistische aanpak.

De consequentie van de probabilistische aanpak is dat benaderingen moeten

worden ingevoerd. Wij hebben gezien dat mean field methodes hierbij van nut kunnen zijn. Wij hebben ons in dit artikel beperkt tot Boltzmann verdelingen en tot het probleem van leren. De mean field methodes zijn echter ook met veel succes op andere kansmodellen en op het probleem van redeneren toegepast.

Over de auteurs

Bert Kappen is UHD bij de vakgroep medische- en biofysica van de Katholieke Universiteit Nijmegen. Hij behaalde zijn doctoraal theoretische natuurkunde in 1983 aan de universiteit van Groningen. Hij promoveerde in 1987 aan de Rockefeller university in New York op het effect van stralingcorrecties aan het standaard model van de zwakke wisselwerking. Van 1987 tot 1989 was hij werkzaam bij het Philips Natuurkundig Laboratorium in Eindhoven op het gebied van de combinatorische optimalisatie ten behoeve van VLSI ontwerp. Vanaf 1989 is hij werkzaam bij de vakgroep Biofysica van de Katholieke Universiteit Nijmegen en bij de Stichting Neurale Netwerken. In 1997 ontving hij een PIONIER subsidie voor zijn onderzoek op het gebied van neurale netwerken. Zijn onderzoeksgroep bestaat momenteel uit ongeveer 10 onderzoekers.

Martijn Leisink is in 1997 afgestudeerd op het onderwerp neurale netwerken bij de vakgroep medische- en biofysica van de Katholieke Universiteit Nijmegen. Hij werkt daar nu als Onderzoeker in Opleiding aan benaderingstechnieken voor Boltzmann Machines en de geldigheid van deze methoden.

References

- [Cox, 1961] Cox, R. (1961). *The algebra of probable inference*. John Hopkins Press.
- [Duda et al., 1976] Duda, R., Hart, P., and Nilsson, N. (1976). Subjective bayesian methods for rule-based inference systems. In *Proceedings National Computer Conference (AFIPS)*.
- [Jaynes, 1983] Jaynes, E. (1983). Where do we stand on maximum entropy. In Rosenkrantz, R., editor, *Papers on probability, statistics and statistical physics*. Reidel publishing company, Dordrecht.
- [Kappen and Rodríguez, 1998] Kappen, H. and Rodríguez, F. (1998). Efficient learning in Boltzmann Machines using linear response theory. *Neural Computation*, 10:1137–1156.
- [Reichenbach, 1949] Reichenbach, H. (1949). *Theory of probability*. University of California Press, Berkeley.

- [Schrödinger, 1947] Schrödinger, E. (1947). The foundation of the theory of probability I,II. *Proceedings of the Royal Irish Academy Series A*, 51:51–66,141–146.
- [Shortliffe and Buchanan, 1975] Shortliffe, E. and Buchanan, B. (1975). A model of inexact reasoning in medicine. *Mathematical Biosciences*, 23:351–379.