

Grondslagen van de Psychologie

college 13
philosophy of mind & cognition
19-03-2009

*Functionalisme Deel II:
- Connectionisme &
- "The Biological Turn in AI"*

recapitulatie

- MR als argument tegen MBIT;
- Functionalisme;
- Functionalisme m.b.t cognitie: De eerste stappen in de AI;
- We hebben gezien: Klassieke AI is niet erg realistisch als model voor de menselijke geest;
- En dus: willen we AI gebruiken om mogelijk iets over onszelf te leren, moeten we daar anders naar kijken.

2

Programma college 13

1. **Connectionisme als alternatief voor klassieke AI:** [1] wat is het?; [2] nadere uitleg a.d.h.v. voorbeelden; [3] voordelen connectionisme; [4] evaluatie;
2. **The biological turn:** Opheffing *strikte* scheiding [a] waarneming, [b] cognitie/bewustzijn en [c] actie/handelen/gedrag – hierbij worden bekende biologische principes in de AI geïntroduceerd.

3

1.1 wat is connectionisme?

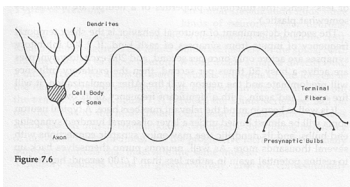
- Dit is een alternatief voor de klassieke seriële, stapsgewijze, regelgeleide kijk op intelligentie & informatieverwerking;
- Andere termen die voor dezelfde kijk gebruikt worden zijn: *Parallel Distributed Processing* (PDP) en (kunstmatige) *neurale netwerk theorie*.

4

TB: 181-186, B. 8.1, G. 271

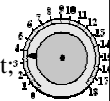
vergelijking biologische en kunstmatige netwerken

- In een biologisch neurale netwerk zijn de eenheden de neuronen.

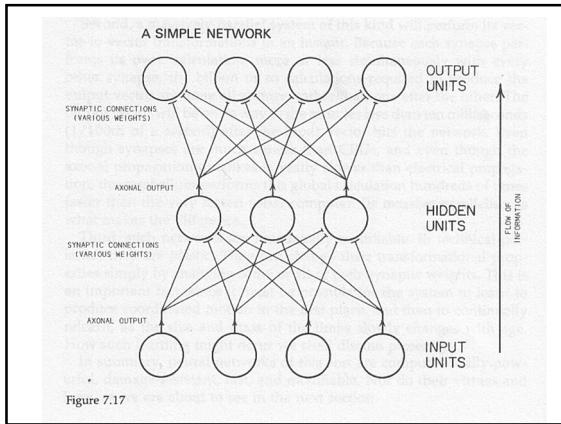


5

- Bij een connectionistisch netwerk heb je om te beginnen *units* en *weights*;
- *Units* zijn de AI "neuronen", ze ontvangen input en verzenden output;
- De *sterkte* van output die een unit verzendt (gerelateerd aan een bepaalde input) is telkens anders: dit noemt met *the weight* (die kan zowel remmend als stimulerend zijn);
- Belangrijk is dat de units niet alleen seriegeschakeld zijn maar ook (voornamelijk) parallel: elke unit kan in principe met elke andere unit verbonden worden.



6



1.2 voorbeelden van connectionistische netwerken

voorbeeld 1: logische neuronen

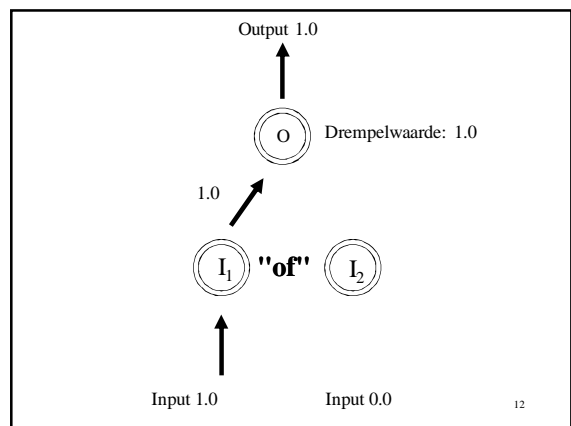
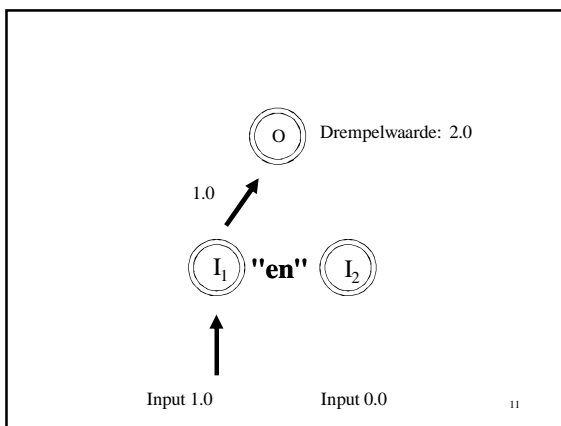
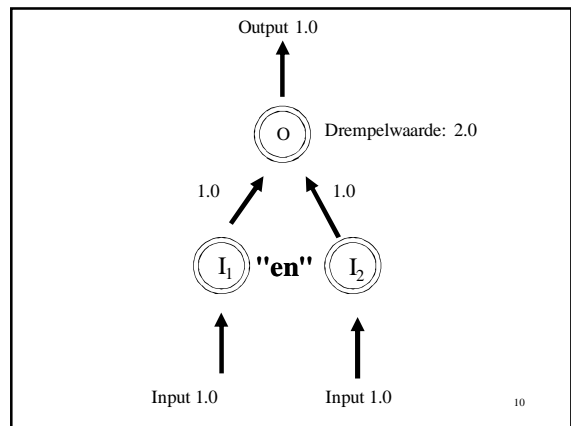
- Het is begonnen met de “logische neuronen” van McCulloch & Pitts (1943);
- Units hebben een bepaalde drempelwaarde;
- Idee: elke functie uit de propositielogica (zoals “en” en “of”) kan je in een PDP systeem maken;
- Bijvoorbeeld “en”.

8

Voorbeeld 1: een eenvoudig netwerkje

- Eenvoudig netwerk: twee input units I1 & I2 en een output unit O.
- De input units geven beide een signaal door naar de output unit, die heeft een drempelwaarde van 2.0.
- O wordt dus pas actief als hij van beide units input krijgt. D.w.z. van I1 & I2.
- Voor “of” kan je hetzelfde doen, je moet dat de drempelwaarde van de output unit verlagen naar 1.0.

9



voorbeeld 2: mijnen opsporen

- **Probleem:** hoe herken je vanuit een onderzeeboot een mijn van een rots? Via radar? Dat werkt niet zo goed.

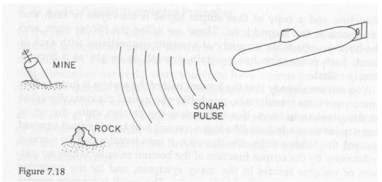
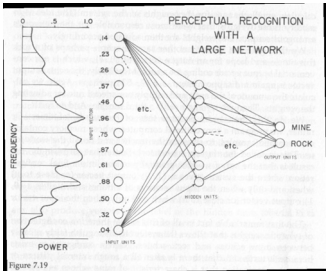


Figure 7.18

13

Stap 1: de sterkte van de frequenties bepalen



PERCEPTUAL RECOGNITION WITH A LARGE NETWORK

Figure 7.19

14

Stap twee: voer dit in een PDP netwerk in

- 13 input units, 7 verborgen units, 2 output units;
- 105 verbindingen;
- In eerste instanties zijn de weights random;
- We willen bij een sonar-input van een mijn, de output $\langle 1, 0 \rangle$. (En voor een rots $\langle 0, 1 \rangle$);
- De output is waarschijnlijk iets als $\langle 0.49, 0.51 \rangle$.

15

Stap drie: het systeem trainen

- Dit gaat via backpropagation;
- Je geeft aan dat de output $\langle 1, 0 \rangle$ moet zijn, en dus moeten de weights – via een algoritme – bijgesteld worden;
- Dat gaat stapje voor stapje (Clark: *gradient descent learning*);

Such learning algorithms can discover solutions we had not imagined.

Clark 2001: 66

16

Voorbeeld 3: gezichtsherkenning

Garrison Cottrell heeft een AI PDP gezichtsherkenningssysteem gemaakt (zie Churchland 1995: 40-55).



Garrison Cottrell

17

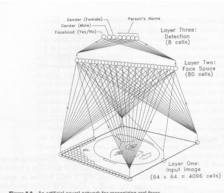


Figure 3.3 An artificial neural network for recognizing half faces.

- (1) **input laag:** een raster van $64 \times 64 = 4096$ cellen;
- (2) **tussenlaag:** waar de *face space* van 80 dimensies zich bevond; 80 cellen dus;
- (3) **detectielaag:** 8 cellen, waarin een onderscheid gemaakt kon worden tussen:
 - $\langle 1, 0, 1, 0.5, 1, 0.5, 0, 0 \rangle$
 - Gezicht.

(a) gezicht/geen gezicht;
 (b) man/vrouw;
 (c) naam van persoon.

18

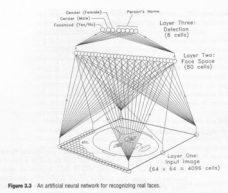


Figure 3.3 An artificial neural network for recognizing real faces.

<1, 0, 1, 0.5, 1, 0.5, 0, 0>
 • Geen man, wel vrouw.

- (1) **input laag**: een raster van $64 \times 64 = 4096$ cellen;
- (2) **tussenlaag**: waar de *face space* van 80 dimensies zich bevond; 80 cellen dus;
- (3) **detectielaag**: 8 cellen, waarin een onderscheid gemaakt kon worden tussen:
 - (a) gezicht/geen gezicht;
 - (b) man/vrouw;
 - (c) naam van persoon.

19

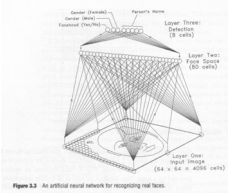


Figure 3.3 An artificial neural network for recognizing real faces.


<1, 0, 1, 0.5, 1, 0.5, 0, 0>
 • Janet.

- (1) **input laag**: een raster van $64 \times 64 = 4096$ cellen;
- (2) **tussenlaag**: waar de *face space* van 80 dimensies zich bevond; 80 cellen dus;
- (3) **detectielaag**: 8 cellen, waarin een onderscheid gemaakt kon worden tussen:
 - (a) gezicht/geen gezicht;
 - (b) man/vrouw;
 - (c) naam van persoon.

20

Door *backpropagation* werden de connecties telkens een beetje bijgesteld als de output niet goed was.

Cottrell's face-recognition network topped out an impressive level of performance. It achieved 100 percent accuracy, on the training set of images, with respect to faceness, gender, and the identity of the face presented.



Paul Churchland

Churchland 1995: 45

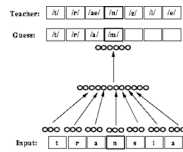
- **Q: Was dit wellicht een kwestie van geheugen?**
- **A: Nee.**
- Ook met andere foto's van dezelfde mensen, herkende het systeem de mensen: 98% accuracy (1 keer zat de computer ernaast);
- Herkende 100% van een set foto's als het ging om gezicht / geen gezicht, en van die gezichten wist het systeem in 81% van de gevallen de juiste sekse te geven.

22

Voorbeeld 4: NETtalk

- NETtalk moest geschreven zinnen omzetten in gesproken woord (of in ieder geval het juiste symbool voor de fonetische klank, dat dan in een spraakcomputer ingevoerd kon worden);
- **DECtalk** deed dat al, op basis van een database met woorden, regels en ingewikkelde grammatuur;
- **NETtalk** is een connectionistisch netwerk zonder zo'n database etc.

23



- Inputlaag bestaat uit 7 groepen van 29 units (in plaatje zie die 7 groepen afgebeeld met 3 bolletjes);
- Ze lezen elk een letter van een woord;
- Het gaat om de middelste letter;
- De rest is context.

24

- Dat is belangrijk, want je moet weten wat de context is: anders zou je “**ghoti**” wel eens als [fish] uit kunnen spreken – GH van tough; O van women; TI van nation;
- Sjenowksi en Rosenberg hebben NETtalk in drie maanden ontwikkeld, en het kan elke taal leren;
- Het werkt ook weer via backpropagation;
- **Backpropagation is biologisch niet erg realistisch**, maar je moet toch ergens mee beginnen.

25

- NETtalk heeft geen regels etc. expliciet opgeslagen;
- Wat heeft het dan wel opgeslagen?
- Het is duidelijk dat de zaak over het hele systeem verspreid zijn opgeslagen, er zijn nogal wat “laesies” nodig om NETtalk zoveel schade toe te brengen dat het niet meer functioneert;
- Onderzoek heeft opgeleverd dat klinkers ergens anders worden opgeslagen dan medeklinkers (dat is ook het eerste verschil dat NETtalk telkens weer leert.)

26

1.3 de voordelen van connectionisme

- 3.a economische manier van representeren;
- 3.b nuances goed aan te geven;
- 3.c schade tolerantie;
- 3.d patroon afmaking;
- 3.e free generalization.

27

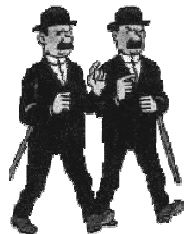
1.3.a economische manier van representeren

- **Opeenstapeling**: je kan dezelfde units en verbindingen gebruiken om vele representaties op te slaan / te activeren (de panter-representatie, de tijger-representatie en het dit-is-de-bakker-op-de-hoek-representatie zitten alledrie in hetzelfde systeem);
- Dit is erg economisch - je hebt niet voor elke representatie nieuwe units (o.i.d.) nodig;
- De representaties zijn **gedistribueerd**.

28

1.3.b nuances goed aan te geven

- Verschillen in aanzicht van 1 persoon / object;
- Of kleine verschillen tussen personen / objecten.



29

1.3.c schade tolerantie

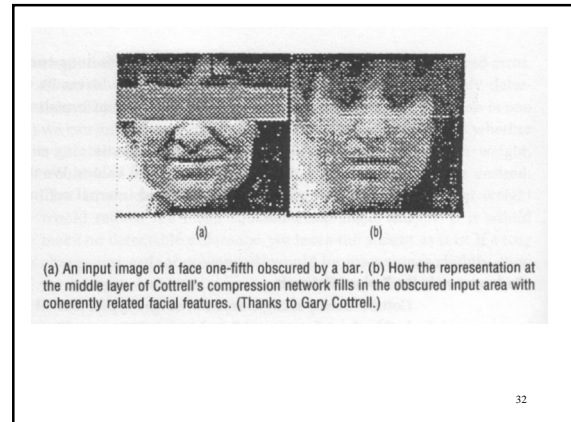
- Doordat het systeem parallel is en niet serieel kan je wel een unit missen en dan toch nog de juiste output verkrijgen, dit uit zich in **graceful degradation**: Het systeem gaat slechts langzaam stuk; het heeft dus een hoge tolerantie voor schade (**damage tolerance**).

30

1.3.d patroon afmaking

- Ook met beperkte informatie komt er nog de goede output uit;
- Met 1/5 van een bekend gezicht kan Cottrells systeem de gezichten nog steeds herkennen;
- Dit gaat echter fout (van 100% naar 71%) als het balkje over het voorhoofd ligt;
- Dit geeft aan dat de haarlijn over het voorhoofd van belang is voor de gezichtsherkenning van dit systeem (Cf. prosopagnosia).

31



32

1.3.e free generalization

- Als je een input A2 (tijger) hebt die sterk lijkt op input A1 (panter, en daar wist je al hoe je erop moest reageren), dan kan je gegeven die input toch vrij adequaat reageren, ondanks dat de situatie nieuw is:

The upshot is that semantically related items are represented by syntactically related (partially overlapping) patterns of activation.

Andy Clark 2001: 66

33

1.4 evaluatie connectionisme

- 1.4.a Levert het systematisch karakter van menselijk denken geen probleem op voor het connectionisme?
- 1.4.b Is het biologisch realistisch?

34

1.4.a systematisch denken

- Clark geeft een eenvoudig probleem dat fataal lijkt te zijn voor connectionisme;
- Het is gebaseerd op het systematisch karakter van ons denken;
- **Systematisch karakter**
 - De Eiffeltoren is groter dan de Martinitoren → De Martinitoren is kleiner dan de Eiffeltoren;
 - Jan gaat op bezoek bij Piet; Piet gaat op bezoek bij Jan.

35

The most famous argument against connectionist models of human thoughts goes like this:
Thought is systematic;
So internal representations are structured;
Connectionist models lack structured internal representations;
So connectionist models are not good models of human thought. (Clark 2001: 76)

36

Reactie Clark

- AC vraagt zich af waar men het idee vandaan haalt dat connectionistische systemen niet met het systematisch karakter van het denken om zou kunnen gaan;
- **HD:** Dat lijkt me inderdaad de beste respons.

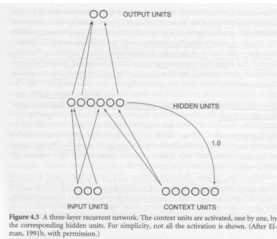
37

1.4.b biologische realiteit

- Een probleem met deze eerste generatie modellen is dat ze enkel aan “snapshot reasoning” kunnen doen: tijd is uit het model gelaten.

38

Recurrent pathways



- Hierdoor ontstaat een soort **korte-termijn geheugen** (maakt actie mogelijk ook als er even geen input is);
- Dynamische systemen: er zijn vele modules die verschillende zaken kunnen (complexer dus).

39

Overeenkomsten met biologie (Churchland 1995)

- Hoge schadetolerantie;
- Snel / real-time;
- Parallel;
- Input gestoord; toch herkenning;
- Free generalisation;
- **HD:** Dit is **inductie**, PDP levert dus een verklaring op voor dat mechanisme in biologische wezens (**Popper had gelijk:** we worden geboren met de verwachting van herhaling[-voor-ons]).

40

2. Een aanpassing van het model

- In zowel het klassieke AI model als het connectionistische is het zo dat er [1] input is, die [2] verwerkt wordt en dan tot [3] output leidt;
- Met betrekking tot mensen zou je dan verwachten dat [1] waarneming de input levert, [2] de hersenen die input verwerken, en [3] de output leveren in de vorm van handelingen & gedrag.....

41

- Dit leidt tot de aanname dat waarneming, cognitie en handelen drie strikt verschillende zaken zijn;
- Andy Clark meent dat dit een te eenvoudige kijk is en dat we ons model zo aan moeten passen dat we dat strikte onderscheid op moeten geven;
- Hij geeft voorbeelden, en kijkt naar de biologie om dit te ondersteunen.

42

2.1 a license to ignore biology

- In de cognitiewetenschap heeft men geprobeerd om cognitie en bewustzijn na te maken zonder naar biologische wezens met bewustzijn en cognitie te kijken;
- Als men naar de biologie gekeken had, dan zou men waarneming, cognitie en handelen niet zo strikt gescheiden hebben als men een lange tijd gedaan heeft.

43

Het is de schuld van David Marr

- David Marr meende dat een verklaring van cognitie uit drie onderdelen (levels) bestaat:
 1. **Taakanalyse:** wat moet het systeem kunnen doen? (bijv. rekenen);
 2. **Representatie & algoritme:** wat voor representaties en regels om die te manipuleren (algoritmes) heb je daarvoor nodig?
 3. **Implementatie:** hoe maak je een machine die dat kan?

44

Als je dit accepteert, dan heb je klaarblijkelijk level 3 niet nodig om iets over level 2 te zeggen: "Until te late 1980s many cognitive scientists took the Marr framework as a license to ignore or downplay the importance of understanding the biological brain." (Clark 2001: 85)

45

2.2 Evolutie: ontwerpen vs. aanmodderen

- Als je wel naar de biologie kijkt, dan kom je al snel bij neodarwiniaanse evolutie-theorie uit;
- Menselijke ontwerpers kunnen aan de tekentafel beginnen om uiteindelijk een ingewikkeld apparaat te maken;
- Maar dat is natuurlijk niet hoe evolutie werkt (er is geen ontwerp / ontwerper).

46

Voorbeeld: evolutie landdieren

- Longen hebben zich ontwikkeld uit de blaas van een vis;
- Evolutie werkt dus met wat ze al heeft & random mutaties die al dan niet in het voordeel zijn (m.b.t. survival & reproduction) van het individu met die mutatie;
- Er zijn dus restricties aan de evolutie, namelijk die opgelegd zijn door wat er al is.

47

- Als mens kom je wellicht met andere oplossingen dan de evolutie;
- Wellicht ben je geneigd om bij het **ontwerpen van een hand** eerst apart de bewegingen van elke vinger te ontwerpen en de zaak dan bij elkaar te brengen als hele hand;
- Evolutie gaat net andersom te werk: de werking van de hele hand (grijpen) is eenvoudiger dan die van een afzonderlijke vinger.

48

- De conclusie is dus dat als we iets na willen maken, we er verstandig aan doen om eerst **eens naar de biologie te kijken** en te zien hoe evolutie via natuurlijke selectie tot een oplossing van het relevante probleem is gekomen;
- Je ziet dan dat waarnemen, cognitie en handelen niet zo radicaal verschillend zijn.

49

Een voorbeeld om dat alvast intuïtief duidelijk te maken

- Het maken van een **legpuzzel**;
- Laten we nu eerst eens kijken naar een aantal biologische principes die nu ook in de cognitiewetenschappen gebruikt worden.

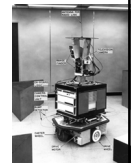
50

2.3 Oude biologische principes in een nieuw AI jasje

- **Principe # 1:** normale interacties tussen een agent en de omgeving zijn **niet afhankelijk van gedetailleerde innerlijke modellen** van de 3D omgeving;
- De gedetailleerde representaties stroken wel met the model van Marr: om iets te laten werken stop je er eerst de representaties en regels in;
- Dit strookt natuurlijk niet met de biologie: er is geen beest dat geboren wordt met een representatie van de omgeving in zijn kop.

Verzet tegen de *filing cabinet* methode

- Vanuit modern onderzoek binnen de cognitiewetenschap komen nu experimenten die aantonen dat onze bewuste representaties van de wereld inderdaad **niet gedetailleerd** zijn;
- (Men trekt dan zelfs de conclusie dat er geen representaties zijn...);
- Wat voor experimenten zijn dat?



52

experimentjes

- Je krijgt om en om twee foto's te zien;
- Foto 1 verschilt op een belangrijk punt van foto 2: wat is het verschil?

53

Haines (1991)

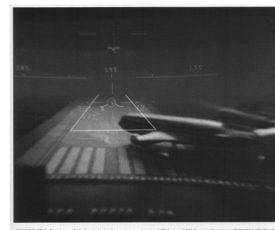


FIGURE 17.4. Simone plea's forward visual scene at an oblique of 75° but not 15° beam with viewing obstructions clearly visible.

54

Principe # 2: Representaties hoeven niet complex / volledig te zijn

- Het is voldoende dat ze adequaat zijn;
- Je hoeft het verschil tussen twee leeuwen niet te representeren, als je leeuwen maar adequaat representeert, d.w.z. als iets gevaarlijks;
- Dat geldt ook voor bewegen in een omgeving;
- Voorbeeld de mobots (mobile robots) van Rodney Brooks....

55

2.4 Evaluatie

- Het is een goede zaak dat cognitiewetenschappers naar de *biologie* gaan kijken en principes daaruit overnemen & gebruiken;
- Dit maakt het model dat ze hanteren biologisch realistischer;
- Conclusie: Functionalisme lijkt te werken voor cognitie / intentionaliteit;
- Vraag: werkt het ook voor fenomenaal bewustzijn / qualia?
- The jury is still out.....

56

Volgende keren

- College 14: hebben we vrije wil?
- College 15: zijn we ons brein?
- College 16: zijn we ons lichaam?

57

Wat voor vragen kan je over deze stof verwachten?

58

Tot volgende week

Vragen?

Mail naar a.j.p.w.dooremalen@uva.nl
Of stel ze op het discussieforum

Inhoudelijke discussie over de stof?

Bekijk het forum op Blackboard

59