

Colline – baggrund og beskrivelse

Resumé

Problemstillingen med nutidens computersystemer ridses hurtigt op, hvorefter de to designstrategier: Blueprints og stage-setting beskrives i relation til kunstig intelligens. Collines virkemåde beskrives herefter overordnet. Som afslutning konkluderes det at Colline som produkt/idé har sine rødder i det nye videnskabelige paradigme der er ved at bryde frem.

Indledning

Nutidens computersystemer registrerer noget i omgivelserne som det herefter reagerer på. Det betyder at systemudviklerne definerer et problemområde og derefter sætter sig ned og prøver at forudse alt hvad systemet vil blive udsat for af brugerne indenfor problemområdet. Dvs. man finder hændelser i problemområdet. Disse hændelser udløser en ”urværksreaktion” i systemet og systemet kommer med en passende reaktion.

På brugeren virker det som om at systemet er intelligent, men i virkeligheden kommer systemets intelligens udefra, nemlig fra systemudviklerne. Det opdager brugeren i det øjeblik at systemet skifter fra at opføre sig ’intelligent’ til at opføre sig direkte tåbeligt. Eksempelvis popper der ustandseligt følgende fejlbesked op på skærmen, idet dette skrives:



Den er tilsyneladende umulig at slippe af med, selv efter flere forsøg på at installere funktionen (som jeg aldrig har bedt om). Den sidste rest af illusion om at systemet er intelligent brister når beskeden, efter at jeg har klikket ’Nej’ 37 gange, endnu engang popper op.

Sagen er at problemområdet for nutidens systemer har vokset ganske eksplosivt i kompleksitet. Systemer skal ikke længere blot udføre en funktion, de skal samarbejde med andre systemer og ændre deres adfærd i forhold til disse systemer.

Ved den sidste virusskanning af min computer, blev der skannet 138.714 filer. Disse filer, som alle er navngivet med hver deres unikke navn, skal fungere som en samarbejdende enhed. Og når der kommer en ny brik til dette gigantiske puslespil, så skal den sættes ind uden at det øvrige korthus falder sammen - ikke underligt at softwareudviklere hører til det mest stressede folkefærd ifølge flere målinger!

Kontrol vs. Autonomi

For at imødegå disse besværligheder har man i flere år forsket i kunstig intelligens¹.

¹ For en kort og præcis gennemgang af kunstig intelligens historien henvises til <http://www.ucla.edu/~isb9112/dept/phil341/wisai/WhatIsAI.html>

Der hvor der er nået de bedste resultater er inden for neurale netværk, men de afgørende resultater/nyskabelser som ville kunne erstatte eller supplere vores nuværende systemer har manglet.

Årsagen til at de afgørende gennembrud har manglet indenfor KI-forskningen, er efter min mening den opdragelse som de der forsker i emnet har fået. De er nemlig blevet opdraget til at bruge deres rationelle og logiske hjernehalvdel og undertrykke den intuitive helhedstænkende hjernehalvdel. Den måde vi kommer frem i undervisningssystemet som det er i dag, er ved hjælp af dyder, som er repræsenteret ved den venstre hjernehalvdel. Om der er noget galt i det skal jeg ikke gøre mig klog på, men det er blot vigtigt at erkende det. For en af følgevirkningerne ved dette er at de programmeringskyndige udviklere som prøver at lave kunstig intelligens, uvægerligt tager een bestemt ting for givet, den er så selvfølgelig for dem og deres kolleger at den ikke engang behøves at blive nævnt: *Kontrol*.

Kontrollen ligger som et baggrundstæppe bagved alle de eksempler på kunstig intelligens jeg er stødt på. Den kan være pakket ind i en matematisk indlæringsalgoritme som kan bevises at føre til den mindst mulige fejl i et neuralt netværk, eller den kan give sig udslag i en kæmpestor database hvor man forsøger at lagre *alt* hvad vi forstår ved sund fornuft.

Det modsatte af kontrol er *autonomi*, og jeg mener at autonomi er det der hidtil har manglet for at få et gennembrud indenfor KI.

Blueprints vs. Stage-setting.

Når man designer systemer med kontrol som baggrundstæppet, så designer man vha. *blueprints*. Det vil sige systemerne bliver analyseret på kryds og tværs i designfasen, så at systemudviklerne har kontrol med alle elementerne.

Hvis man designer med autonomi som baggrundstæppe designer man vha. *stage-setting*. Stage-setting vil sige at man skaber nogen spilleregler og nogen spillebrikker ("legoklodser") og så sætter man ellers blot systemet i gang, evt. puffer man det i den rigtige retning hvis det kommer lidt ud af kurs.

Der er tre betingelser for at stage-setting strategien skal kunne bruges:

1. der skal være et programmeringssprog som skal kunne simulere det der foregår 'på scenen'
2. computerkraften skal være stor nok til at aktiviteten på scenen kan simuleres på en rimelig tid.
3. de legoklodser der 'leger' på scenen skal have de rette egenskaber så de kan sætte sig sammen til en maskine, eller rettere organisme, der løser opgaven.

Da det objektorienterede programmeringssprog gjorde sit indtog i softwareverdenen blev betingelse 1 opfyldt og betingelse 2 er også så godt som opfyldt. Den sidste betingelse er jeg efter 4 års arbejde, on and off, også ret sikker på er opfyldt.

Colline

Forestil dig at man i stedet for at lave en meget kompleks processor laver en processor (agent) der kan følgende:

- Købe en bitstreng, fx [1000011]
- Lede efter "01" i denne bitstreng og spise den hvis den er der
- Hvis "01" var i bitstrengen sættes agentens tilstand til "1" (sand) ellers "0" (falsk)
- Sælge tilstanden til interesserede købere
- Sælge den resterende bitstreng videre

Og ydermere:

- parre sig med en anden agent når fitness er vokset til det dobbelte
- bevæge sig rundt for at finde bitstrengene eller egnede agenter til at parre sig med

Denne type agent kaldes inpoder. Der laves to varianter af denne agent, collector og effector, så der i alt er tre forskellige typer agenter. Kort sagt så er inpoderens opgave at processere beskeder, collectoren samler små beskeder til en større og effectoren bestemmer sin tilstand ud fra to andre agents tilstand. De tre typer agenter repræsenterer tilsammen en helhed, dvs. de kan sættes sammen ligesom legoklodser, på en måde så at alle opgaver indenfor mønstergenkendelse principielt kan løses.

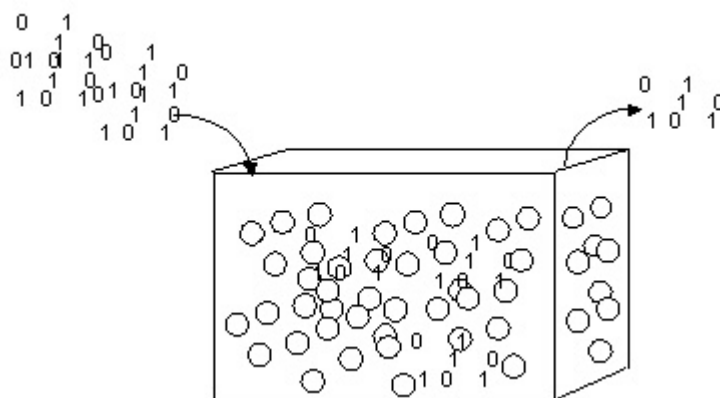
Nogle tusinde af disse agenter placeres i et 3D netværk og hver af dem får herefter følgende besked:

Du lever i 3 sekunder, på den tid skal du nå at:

1. fordoble din fitness ved at købe og sælge beskeder
2. finde en anden agent at parre dig med inden du dør

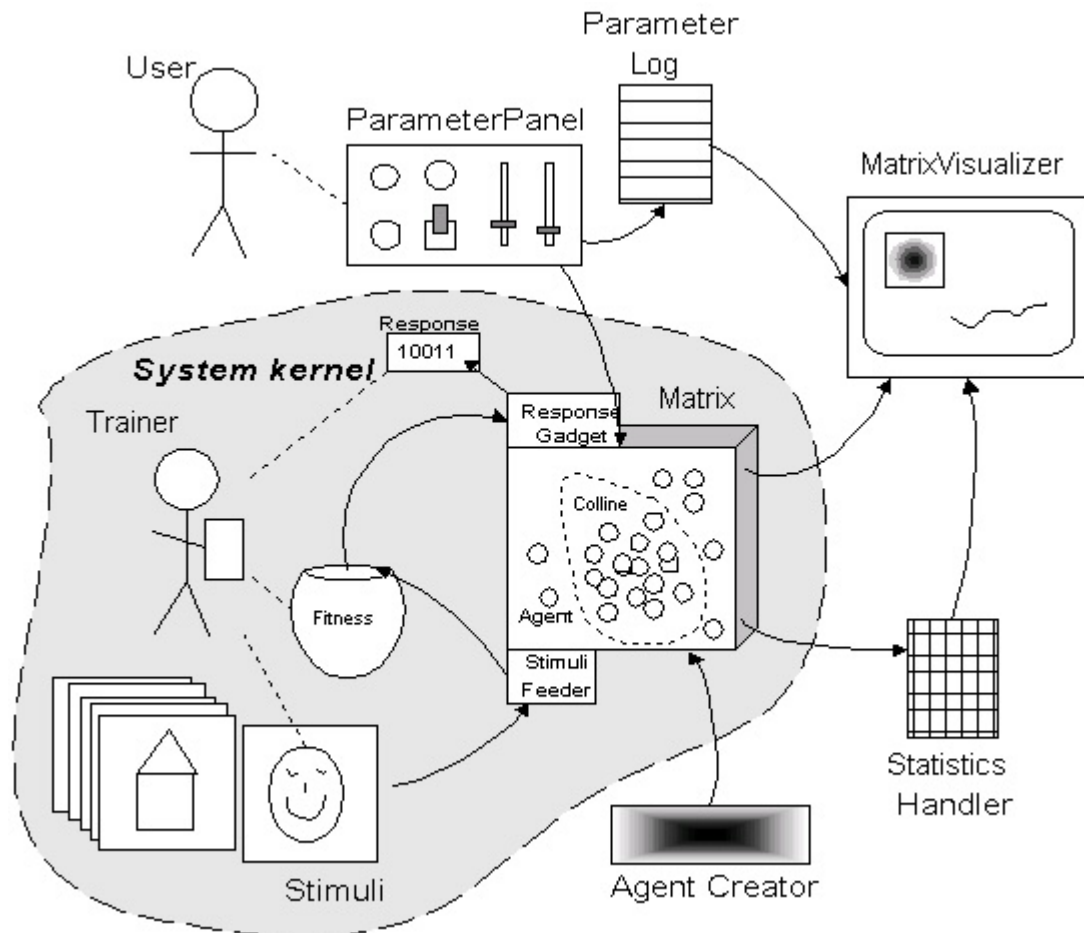
Start!

Herefter begynder træneren at præsentere information fra omverdenen som agenterne kan vælge at reagere på (købe). I den anden ende af "akvariet" indsamles et respons fra en gruppe udvalgte agenter. Princippet er illustreret på figur 1.



Figur 1: Colline kan betragtes som et akvarium med små agenter, der handler med beskeder.

Et mere illustrativt billede af systemets virkemåde er vist på figur 2.



Figur 2: Rigtigt billede af Colline og arbejdsmiljøet.

Hvad sker der så?

Jeg har lavet en pilot implementering i C++, med 1-3000 agenter i et 2D netværk og resultatet er yderst interessant:

I starten uddøde agenterne efter kort tid, men efter nogle forsøg fik jeg indstillet parametrene² så der sker to ting, nemlig:

1. Bestanden af agenter stabiliserer sig på et bestemt niveau
2. Outputtet begynder at korrelere med inputtet

Eller sagt på en anden måde:

1. **Systemet bliver stabilt**
2. **...og begynder at udvise intelligens**

De svar som systemet giver vil i starten være fuldstændig tilfældige, men eftersom træneren belønner de agenter som svarer rigtigt og ignorerer de andre, begynder det – efter nogle generationer - at påvirke agenternes adfærd.

² fx agentlevetid, hvor stor procentdel af fitness forældre overfører til børn, mutationsrate, m.fl.

Jeg mener der her er en ret præcis analogi til den måde børn opdrages på:

”I en berømt undersøgelse af vuggestuebørn i det første leveår, opdagede den amerikanske psykoanalytiker Margaret Mahler, at børnene udsendte en mangfoldighed af signaler og tegn for at gøre opmærksom på behov og følelser af forskellig art, men at moderen var meget udvælgende i sin reaktion på disse tegn. Som et resultat af dette ændrede barnet gradvis sin adfærd for at tilpasse sig moderen.”

[Vedfeldt, 'Bevidsthed – bevidsthedens niveauer']

De agenter der får opmærksomhed (dvs. fitness point) blomstrer, og de sætter deres præg på hele agentkulturens opførsel. Agenterne begynder enkeltvis at tilpasse sig (via generationsskift) hinanden i en slags 'myretue-dans' hvor de gradvist lærer hvilke agenter de skal kigge efter og hvordan de skal reagere på bestemte signaler³. På den måde opstår der en korrelation imellem inputtet og outputtet, bestemt af træneren. Det kan f.eks. være genkendelse af et bestemt mønster. Korrelationen opstår fordi der begynder at opstå samarbejds mønstre imellem agenterne, eller *vaner* med et andet ord. Om cellerne i vores krop skriver Jesper Hoffmeyer:

”Cellens 'intelligens' beror udelukkende på vaner nedlagt gennem dens milliardårige historie.”

[Hoffmeier, 'En snegl på vejen']

Indlæring

Disse 'vanedannelser' kan studeres i Colline og ved at skrue på parametrene kan det bestemmes hvor hurtigt agenterne skal krystallisere i et bestemt samarbejds mønster. Når et samarbejds mønster er opstået, betyder det at systemet har løst opgaven så godt som det evner med de aktuelle parameterindstillinger. Systemet svarer f.eks. 'ja' til 90% af de cirkler det præsenteres for. Hvis trænerens krav pludselig ændres, f.eks. kræver han at der kun svares 'ja' hvis der er en prik inde i cirklen, så opstår der en 'spænding' i systemet. Den nødvendige fitness ledes nu ikke længere ind i det vanemæssige samarbejds mønster. Dette giver et fitness overskud hos agenterne udenfor det dominerende skema/ samarbejds mønster, hvilket giver grobund for at rebellerne i systemet bliver hørt, dvs. de agenter som processerer beskeder og samarbejder på en anden måde end det dominerende skema.

Rebellerne suger dette fitnessoverskud til sig og begynder at danne et alternativt samarbejds mønster. Der opstår herefter en konkurrence imellem disse to samtidige samarbejds mønstre. Hvis det nye samarbejds mønster løser opgaven bedre end det gamle (det genkender prikkerne i cirklerne) vil dette mønster udkonkurrere det gamle mønster.

Man kan også sige at det nye samarbejds mønster medfører færre spændinger end det gamle⁴. Systemet har lært noget nyt – det er blevet mere intelligent.

³ ”Korrelationer i adfærd implicerer en slags fælles kode, eller protokol, ved hvilken tilstandsændringer i en celle kan genkendes og forstås af en anden som et meningsfuldt signal. Uden korrelationer i adfærd kan der ikke være nogen fælles kode at kommunikere information med” [Chris Langton]

⁴ ”(Harmoniteorien) Når et netværk arbejder, vil det søge at tilfredsstille flest mulige restriktioner eller 'constraints' og især de stærkeste af dem. Det kaldes derfor et constrainttilfredsstillende netværk. Ved opgaveløsning tilstræber netværket så lavt et samlet energiforbrug som muligt. Dets mål er 'frustrationsformindskelse' og 'afspændning' (Rumelhart, D. et alii, 1986, s. 9-13)./ En svag restriktion svarer til en lille energetisk ladning, en stærk restriktion svarer til en stor energetisk ladning. Ved input svinger systemet i begyndelsen mellem forskellige fortolkningsmuligheder, men på et tidspunkt låser det sig fast på een mulighed, som hermed bemægtiger sig energi fra konkurrerende fortolkningsmodeller./ Netværkets funktion forklares ud fra en 'harmoniteori', som hedder således, fordi der ved fortolkning aktiveres skemaer, som søger den mest 'harmoniske', dvs. sandsynlige tilstand i

Enhederne i Colline har en tilstand og en adfærd, i modsætning til neurale netværk, hvor enhederne er stationære og løser en matematisk funktion. Denne tilstand og adfærd er ydermere karakteriseret ved kreativitet, idet agentens børn ikke er en kloning af agenten. Dette betyder at problemløsningen og indlæring *sker på systemets eget initiativ*, systemudvikleren har blot opsat spillereglerne. At det er lykket at få et system til at udvise initiativ er så vidt jeg ved en ny ting indenfor kunstig intelligens forskningen.

Indlæringen, eller organiseringen i mere komplekse strukturer, kan sammenlignes med den måde vores bevidsthed udvikler sig:

”Udviklingen af personligheden sker i sekvenser fra primitivt til mere differentieret, fra lavere til højere kompleksitet. Overgangen fra en mindre til en mere udviklet fase indebærer, at et højere niveau styringssystem kommer til syne i en kreativ proces, og at det bliver besat med energi, som tidligere var bundet på lavere systemniveauer.”

[Vedfeldt, Bevidsthed – bevidsthedens niveauer]

Begrebet ’energi’ svarer til fitnesspoint i Colline.

Paradigmevalg

Der hvor andre forsøg på at efterligne menneskelig intelligens har fejlet, er efter min mening, at forsøgene hænger fast i en traditionel analytisk og logisk fremgangsmåde, som dræber kreativiteten:

”Hvad jeg opponerer mod er, at man vil presse livsprocesserne ind i den slags matematik, hvor der står noget på hver side af et lighedstegn. For det indebærer, at det på højre side ’ikke er andet end’ det på venstre side, hvilket kun er sandt, hvis man har pillet kreativiteten ud af systemet, dvs. slået det ihjel.”

[Hoffmeyer, ’En snegl på vejen’]

De to tilgangsvinkler til løsning af KI-problemet, beskrives af Ole Vedfeldt:

”Både i filosofien og kunstig-intelligensforskningen har der været to grundlæggende forskellige opfattelser af, hvordan vor ganske almindelige hverdagsviden og praktiske færdigheder bliver til. Den ene opfattelse hævder, at det sker ved logiske følgeslutninger ud fra regler, den anden, at det sker på en erfaringsmæssig, intuitiv og ikke-analytisk måde.”

[Vedfeldt, ’Ubevidst Intelligens’]

Colline er, så vidt jeg er orienteret, det første eksempel på kunstig intelligens som helt har sluppet ”navlestrengen” i den logisk/analytiske tilgangsvinkel, og helhjertet bruger den intuitive indlæringsmåde. Colline som produkt/idé suger sin næring fra det nye videnskabelige paradigme som er ved at bryde frem⁵.

Ifølge Habermas kan videnskaben opdeles i tre forskellige grupper med hver deres fokus:

omgivelserne. Systemet starter ved ’høj temperatur’, det udfører en søgeproces begyndende med kombinationer af lokale løsninger, men køler så ned for at ende med en global løsning, som det ’fryses fast’ på. Det går fra en flydende til en mere fast tilstand. Under denne proces foregår der mange udvekslinger frem og tilbage mellem enhederne, der gensidigt justerer hinanden, indtil det mønster er fundet, som har maksimal harmoni, dvs. passer bedst med omgivelserne.” [Vedfeldt]

⁵ For yderligere info om hvordan det nye paradigme vil påvirke softwareudviklingen, henvises til artiklen på <http://polaris.ing.unimo.it/Zambonelli/PDF/KER.pdf>

”Filosoffen Jürgen Habermas udviklede i 1960'erne en erkendelsesteori om tre forskellige typer af videnskaber med hver deres 'erkendelsesinteresse': naturvidenskaben, hermeneutikken og de kritiske videnskaber.

Naturvidenskabernes forskningsmotiv er teknisk. De vil kontrollere objektive processer og skabe information om regelmæssigheder i naturen. De hermeneutiske videnskaber søger at skabe livsmuligheder, selvforståelse og forståelse af andre. De interesserer sig for betydning og mening. De kritiske videnskaber har som grundlæggende motiv menneskets frigørelse fra historiske og samfundsmæssige kræfter. De søger at regulere livsførelse, holdninger og mål i tilværelsen. De kræver enkeltindivider, som kan hæve sig op over gruppenormer, og forudsætter intellektuel virksomhed, som kan frigøre bevidstheden fra fordomme og forforståelser.”

[Vedfældt, 'Ubevidst intelligens']

Det nye videnskabelige paradigme er opstået som en syntese af disse tre områder, fordi forskere indenfor alle forskningsgrene har opdaget at det bliver mere og mere nødvendigt at tænke på tværs af de forskellige discipliner. Hertil kommer et mere og mere påtrængende krav fra samfundet generelt om *helhedstænkning*.

For yderligere info om det nye paradigme, se pdf'en 'Scientific revolution'.

For yderligere info om de forretningsmæssige anvendelsesmuligheder for Colline, se pdf'en 'Colline – anvendelsesområder'.