

SPROGMUSEET

Redaktør: Ole Stig Andersen



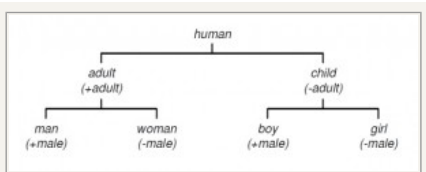
Fra Wittgenstein til Ontology Engineering 3: Kognitive mønstre og neurale netværk

Af Erik David Johnson 11. maj 2012 • I kategorien [Sprogteknologi](#) •   

I den [anden artikel i denne serie](#) fastslog vi hvorledes den kognitive semantik, eksemplificeret ved Lakoffs afdækning af sprogets underliggende, metaforiske struktur, udviser potentiale for at nå frem til en systematik i sproget som afspejler vores fælles livsform. Samtidig så vi også hvordan repræsentationsformen for denne kognitive semantik er afgørende for at det kan lykkes fremtidens forskere i den sprogteknologiske side af kunstig intelligens at udnytte denne systematik til at udstyre computere med egentlig 'world knowledge'.

World knowledge som ontologi

Som nævnt tidligere er world knowledge den viden om verden som sprogbrugere skal have for at kunne opnå en succesfuld sprogbrug. Hvis en computer skal kunne håndtere world knowledge er den altså nødt til at have en form for vidensbase, som kan indeholde informationer om verdens struktur. Med andre ord, så er den nødt til at implementere en *ontologi*. Som en filosofisk term kan en ontologi i den bredeste forstand siges at være det der eksisterer i verden. Indenfor den sprogteknologiske side af kunstig intelligens, har ontologier typisk en hierarkisk struktur af underklasser af forskellige ting og begreber, og typisk indenfor et meget begrænset domæne.



Her ses et simpelt eksempel på en ontologi for et ekstremt begrænset domæne, bestående af hierarkisk strukturerede kategoriseringer.

Arbejdet med udvikling af ontologier, der har fagtermen '[Ontology Engineering](#)', er endnu et meget ungt felt, som stadig er gennemsyret af akademisk rivalisering og uenighed om de forskellige, divergerende tilgange til implementering, graden af formalisme og i særdeleshed det *konceptuelle spektrum*. De fleste ontologier beskæftiger sig dog med et begrænset domæne, som for eksempel de begreber der er relevante for faciliteringen af sprogligt assisteret booking af en flybillet, og som regel er disse ordnet i et simpelt system af klassehierarkier.

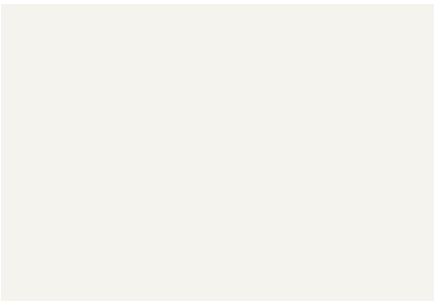
Sådanne ontologier minder i høj grad om den type ontologi man finder hos den *tidlige* Wittgenstein, og det skulle gerne stå klart at de således ikke formår at understøtte et natursprogs kontekstuelle *brug*, ved at inkorporere strukturer der repræsenterer vores fælles livsform i form af diverse kognitivt semantiske mønstre. Pointen er at en samlet ontologi der kan indeholde world knowledge ud fra den tidligere etablerede sprogfilosofiske og kognitivt semantiske præmis, bliver nødt til at tilgå feltet 'Ontology Engineering' på en helt ny måde. Det er utilstrækkeligt selv at kortlægge en model over forskellige identificerbare klassifikationer. I stedet må vi forsøge at udvikle en beregnet model der vil kunne trænes til at indeholde den systematik, der tidligere viste sig at være repræsenteret i natursprog i form af kognitivt semantiske mønstre indeholdt i vores bevidsthed i kraft af vores fælles livsform.

Neurale netværk som beregnet model

Den menneskelige, biologiske hjerne består bl.a. af nerveceller, der kaldes neuroner. Disse neuroner danner et såkaldt neuralt netværk. Mellem neuronerne i dette netværk sker der en udveksling af elektrokemiske signaler. En neuron får input fra andre neuroner og sender selv et output som resultat ud på netværket. Når vi lærer forstærkes forskellige forbindelser, og det er groft sagt måden hvorpå den menneskelige hjerne registrerer og holder forskellige typer af informationer og viden.

Indenfor kunstig intelligens arbejder man med en beregnet model som fungerer ud fra samme principper og som derved kan ses som en primitiv simulation af den menneskelige hjerne. Modellen kaldes for et 'kunstigt [neuralt netværk](#)'. Styrken ved neurale netværk er, at de ud fra en række eksempler er i stand til at lære mønstret i den viden der er repræsenteret ved de givne eksempler og derefter anvende denne viden på nye eksempler der ikke er set tidligere. De er, så at sige, i stand til at *generalisere på data*. Dette virker måske lidt verdensfjernt så lad os se på et simpelt, tænkt eksempel.

Lad os forestille os at vi har en stor database, som indeholder et sort/hvidt billede af hver eneste danskers ansigt. Vi ønsker at opdele databasen i mænd og kvinder ud fra deres udseende. At kategorisere samtlige billeder manuelt ville tage en evighed. I stedet føder vi et kunstigt neuralt netværk nogle eksempler på mande- og kvinde ansigter. Hvert billede den får laves om til en række binære værdier som er enten 0 eller 1 – det kunne for eksempel være hver pixel i billedet som enten er sort eller hvid – og vi angiver også den korrekte



Seneste sprognyheder

4/5	Sprogforskerne fandt en skat i skoven politiken.dk
27/4	Lille indsats styrker små børns sprog www.dr.dk
26/4	Lad os komme det danske 'jantekomma' til livs politiken.dk
23/4	Ud med sproget - Berlingske Mener www.b.dk
20/4	Unikt runefund i centrum af Odense videnskab.dk

4/1	John Holm, Pioneer in Linguistics, Dies at 72 www.nytimes.com
10/8	Young women, give up the vocal fry and reclaim your strong female voice www.theguardian.com
20/5	Bill Funding Native Language Programs Passes mtpr.org
17/5	Sounds Of The Pilbara II: Songs In Language finishes recording - WAM - West Australian Music wam.org.au
13/5	Seven US Senators Introduce Bill to Promote Preservation of American Indian Languages - Native News Online nativenewsonline.net

FLERE NYHEDER >>>

Verdens sprog på Sproguseet.dk på et større kort

Mere i kategorien 'Sprogteknologi'

Fra Wittgenstein til Ontology Engineering 2: Kropslighedstenen og den kognitive semantik

Fra Wittgenstein til Ontology Engineering 1: Analytisk Sprogfilosofi og Moderne Kunstig Intelligens

Nye kommentarer

Arturo til Hebraisk: Et genoplivet sprog eller et nyt sprog?

Yunus til Two Turkish Loanwords in Swedish

Sonstige til Den Danske Ordbog på nettet

Mads Haupt til Er det forkert at splitte sammensatte ord?

Monica Scheuer til Jødiske efternavne

jane til Jødiske efternavne

Inglês til Sprogene i Mozambique

Birgit Eggert til Hvad er der blevet af Maren?

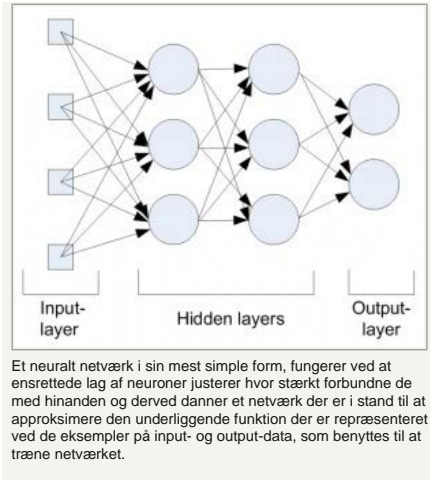
Grethe Movsing til Hvad er der blevet af Maren?

Henrik Klindt-Jensen til Ded borrihjolmska måled

Artikler om

kategorisering for hvert eksempel. Det kunstige neurale netværk vil i teorien kunne lære den underliggende funktion der bestemmer en persons køn ud fra deres ansigt, og anvende den til at kategorisere de resterende billeder.

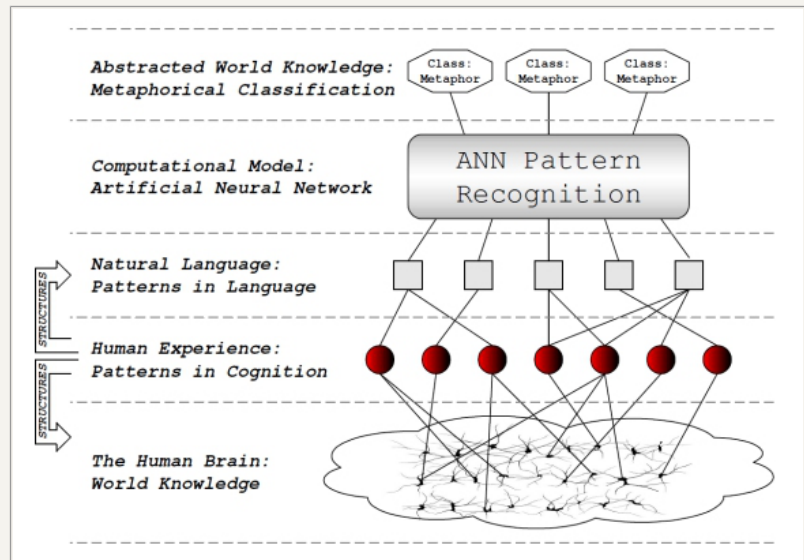
En vigtig pointe ved kunstige neurale netværk er dog, at uanset hvor god en repræsentationsform for data man finder, og uanset hvor mange gode eksempler der benyttes til at træne det, så vil det trænede netværk, selv i bedste tilfælde være en *approximering* af den underliggende funktion det skal repræsentere. Til gengæld kan et neuralt netværk repræsentere meget komplekse og ukendte funktioner, så længe der findes korrekte eksempler som kan bruges til at lære fra. Hvis der er et mønster i de eksempler der bruges til at træne det, så vil et kunstigt neuralt netværk altså kunne finde det, og anvende det på nye data.



Mønstergenkendelse af sprogets kognitive semantik

Den opmærksomme læser vil måske allerede nu have gjort en kobling mellem de tidligere beskrevne metaforisk orienterede strukturerer i den menneskelige hjerne og kunstige neurale netværks potentiale for at kunne simulere hvordan den menneskelige hjerne fungerer. Jeg må dog på det kraftigste advare imod denne ellers meget nærliggende tankegang. Hvis vi fastholder hypotesen et øjeblik kunne vi forestille os hvorledes et neuralt netværk kunne tage et input der repræsenterede vertikalitet og et andet der repræsenterede mængde, og derved repræsentere metaforen [Mere er op; Mindre er ned]. Men hvad kan en sådan en simpel konstruktion fortælle os om måden hvorpå vores fælles livsform udmønter sig i en kognitivt semantisk systematik i sproget? Svaret er: Ikke særligt meget.

I stedet skal kunstige neurale netværk anvendes til det de gør bedst – mønstergenkendelse. Hvis vi kan finde en repræsentationsform af sprogets kognitive semantik der er forenelig med den type input som et neuralt netværk fordrer, så vil vi ud fra sproglige eksempler kunne opbygge en ny type ontologi, som ud fra trænede neurale netværk ville kunne tolke sproglige konstruktioner samt deres betydning, ud fra deres relation til vores fælles livsform. Man kunne for eksempel forestille sig at et kunstigt neuralt netværk kunne trænes til at identificere eksempler på metaforiske koncepter, og derved aflæse en menneskelig samtalepartners idiomatiske billedsprog. At "føle sig lidt nede" har intet med ens fysiske placering at gøre, og det skal fremtidens sprogteknologiske systemer kunne identificere.



I teorien kan vi, vha. værktøjer til mønstergenkendelse såsom kunstige neurale netværk, approksimere de mønstre der konkret er udmøntet ved neurale forbindelser i den menneskelige hjerne, fordi de igen udmønter sig direkte i kognitive mønstre, som vi deler i kraft af vores fælles livsform, og som derfor også kommer til udtryk på en ensartet måde i vores sprog. Derved kan et nøgleelement i simuleringen af den menneskelige bevidsthed siges allerede at ligge i vores sprog og dets evolution.

Hvor skal vi hen?

Vi har i de to tidligere artikler etableret nødvendigheden af en ny tilgang til den sprogteknologiske side af forskning i kunstig intelligens, som i stedet for at have en statistisk tilgang til de nederste niveauer og sproganalyse, starter ud med det øverste niveau af sproganalyse for successivt at lade det definere de nedre niveauer. Vi skal altså tage

aktuelle sprog Alfabeter Anmeldelser arabisk

Biblen bogstaver børn Danmark Dansk Dialekter

engelsk esperanto Formidling fransk identitet

konsonanter Medier modersmål Musik Navne norsk Ord

ordbøger ordforråd oversættelse Plansprog religion

romanske sprog russisk Sjøv skriftsprog sprogdød Sproggeografi

sprogkort Sprogpolitik sprogteknologi

svensk truede sprog tv tyrkisk tysk Udtale

Underholdning video vokaler

Arkiv	Resources
januar 2015	Ethnologue: Languages of the World
december 2014	Forvo – All the Words in the World. Pronounced.
november 2014	LL-Map: Language and Location
maj 2014	Minority Rights Group
marts 2014	OmniGlot. Writing Systems and Languages of the World
februar 2014	UNESCO Atlas of the World's Languages in Danger
oktober 2013	World Atlas of Linguistic Structures (WALS)
august 2013	
marts 2013	
januar 2013	
december 2012	
november 2012	
oktober 2012	
september 2012	
juli 2012	Bogstavlyd
juni 2012	Dansk sprognavn
maj 2012	Den danske ordbog
april 2012	Dialekt.dk
marts 2012	dk.kultur.sprog
februar 2012	Korpus.dk
januar 2012	Nye ord i dansk på nettet (NOID)
december 2011	Ordbog over det danske sprog
november 2011	Ordnet. Dansk sprog i ordbøger og korpus
oktober 2011	Sproget.dk
september 2011	Svenska Akademien
august 2011	Θ (Schwa.dk)
juli 2011	
juni 2011	
maj 2011	
april 2011	
marts 2011	
februar 2011	
januar 2011	
december 2010	
november 2010	
oktober 2010	
september 2010	
juni 2010	
maj 2010	
april 2010	
marts 2010	
februar 2010	

udgangspunkt i world knowledge. Vi så også hvorledes denne vidensbase bør være orienteret omkring vores fælles livsform og måden hvorpå den udmønter sig i sproget og vores bevidsthed, i form af en kognitivt semantisk systematik.



Erik David Johnson er en nyligt udklækket cand.it med speciale i kunstig intelligens og sprogteknologi fra IT-Universitet i København. Hans speciale, som lægger op til en helt ny og hidtil uprøvet tilgang til forskning i den sprogteknologiske side af kunstig intelligens, tiltrak sig for nylig opmærksomhed, da det viste sig at dets centrale indhold og pointe svarede fuldstændigt til det teoretiske udgangspunkt for et endnu ikke offentliggjort 5-års forskningsprojekt på Berkeley Universitet, som begynder i 2012.

Et godt eksempel på en sådan systematik er sprogets underliggende, metaforiske strukturer, men jeg vil understrege at det er muligt at denne ikke kan udgøre det hele billede. Hvilke former for kognitivt semantisk systematik der kan anvendes til at udvikle en ontologi der afspejler vores fælles livsform, er netop et fokus som bør tages op af fremtidig forskning i den sprogteknologiske side af kunstig intelligens. Ligeledes er kunstige neurale netværk ikke den eneste metode til at skabe en beregnet model der kan trænes til at håndtere en sådan systematik. Dette bør også være et fokus for denne type af forskning, og selv hvis en sådan ontologi kan findes, så er det meget muligt at den vil skulle komplimenteres af en traditionel hierarkisk opbygget ontologi i en kompleks arkitektur der kan udgøre den endelige vidensbase for fremtiden natursprogssystemer.

Under udviklingen af disse ideer fandtes der endnu ikke konkrete eksempler på denne type forskning, men gennem korrespondance med George Lakoff selv, blev forfatteren til denne serie af artikler gjort opmærksom på at afdelingen for forskning i kunstig intelligens på Berkeley Universitet i Californien, USA, som beskæftiger sig med den mest avancerede type af forskning i sprogteknologi, lige havde fået en 5-års bevilling af den amerikanske regering til netop at forske i udviklingen af en ontologi, der tog udgangspunkt i sprogets kognitivt semantiske systematik, sådan som det forsås her. Projektet er i skrivende stund endnu ikke offentliggjort, men dets udgangspunkt validerer i høj grad de beskrevne teser og teorier, og det bliver således ikke helt uinteressant at følge med i hvilke forskningsmæssige resultater vi vil se fra Berkeley Universitet i løbet af de kommende år.

Erik David Johnson

www.erk.dk

januar 2010
december 2009
november 2009
oktober 2009
september 2009
august 2009
juli 2009
juni 2009
maj 2009
april 2009
marts 2009

Læs også:

1. [Fra Wittgenstein til Ontology Engineering 2: Kropslighedstesen og den kognitive semantik](#) George Lakoff er en fremtrædende figur indenfor kognitiv semantik, og betragtes af mange som faderen til kropslighedstesen. Udover sine analyser af politisk billedsprog, er han især også kendt for sit...
2. [Fra Wittgenstein til Ontology Engineering 1: Analytisk Sprogfilosofi og Moderne Kunstig Intelligens](#) Hvis forskere indenfor de sprogteknologiske aspekter af kunstig intelligens skal tage ved lære af Wittgensteins voldsomme revidering af sin egen sprogforståelse, så må vi først og fremmest forkaste alle afarter...
3. [Ja, kreolsprog er anderledes](#) Jeg skrev for nylig om kreolsprog. Mikael Parkvall (Stockholms Universitet) havde bevist at kreolsprog var blandt de mindst komplekse i verden, på basis af en omfattende gennemgang af næsten 200...
4. [Hvad sker der med dansk skriftsprog på internettet?](#) En stadig større del af de tekster som offentliggøres og læses på internettet, er skrevet af ikkeprofessionelle skribenter, og tekster af professionelle og ikkeprofessionelle optræder mellem hinanden. I forhold til...

Tagget med: Lakoff, metafor, mønstergenkendelse, neurale netværk, ontologi, sprogteknologi, Wittgenstein, world knowledge

Skriv en kommentar

Navn (kræves)

E-mail (kræves)

Hjemmeside

Send mig en e-mail når der kommer flere kommentarer.