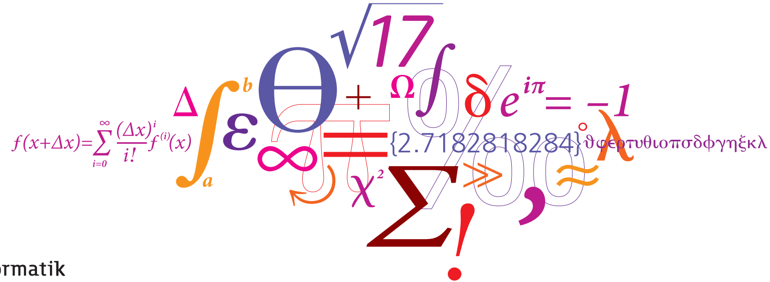
Epistemisk logik og kunstig intelligens

Thomas Bolander, DTU Informatik

Gæsteforelæsning i
Kognitionsforskning II, CST, KU, efteråret 2009



DTU Informatik
Institut for Informatik og Matematisk Modellering

Logik

Logik er et meget bredt interdisciplinært emne med forbindelser til matematik, filosofi, datalogi (herunder kunstig intelligens), lingvistik og kognitionsvidenskab.

Logik handler essentielt om to ting:

1. Hvordan man formaliserer udsagn og giver dem en matematisk præcis betydning.
2. Hvordan man foretager formel ræssonering med disse udsagn.

I filosofi er disse ting interessante, da logik kan benyttes til at (forsøge at) give præcis betydning til begreber som vi ellers kun har en vag definition af, såsom viden, tro, sandhed, pligter, identitet, eksistens, nødvendighed m.m. Studiet af at præcisere disse begreber indenfor en logisk ramme kaldes **filosofisk logik**.

I dette foredrag vil jeg fokusere på begreberne viden og tro, og hvordan disse forsøges at indfanges logisk.

Epistemisk logik

Epistemisk logik er betegnelsen for logiske systemer hvori man forsøger at formalisere udsagn vedrørende viden, såsom "**jeg ved** at jeg nu er på det nye KUA", "**jeg ved** at to plus to er fire", "**jeg ved ikke** om alle kursets deltagere er tilstede i dag".

Doxastisk logik er betegnelsen for logiske systemer hvori man i stedet forsøger at formalisere udsagn vedrørende tro/overbevisning, såsom "**jeg tror** at boligpriserne vil fortsætte med at falde".

Ofte bruges betegnelsen *epistemisk logik* til at dække bege dele, som vi også vil gøre det her.

Det kan i det hele taget være vanskeligt at skelne skarpt og formelt mellem tro og viden. Hvis min søn siger "syv plus syv er femten" siger jeg "min søn **tror** syv plus syv er femten". Hvis han siger "syv plus syv er fjorten" siger jeg "min søn **ved** syv plus syv er fjorten". Men der vil nok næppe være forskel i hans mentale repræsentation af de to udsagn.

De mudrede børn

For at finde ud af hvordan man bedst kan formalisere viden i en logisk ramme kan det være en fordel at se på nogle konkrete eksempler på situationer som involverer viden. En berømt gåde involverende viden er **de mudrede børn** (muddy children) (Barwise, 1989). Gåden forløber på følgende måde:

> *Nogle børn som har været ude at lege kommer ind og deres far siger: "Mindst én af jer har mudder i panden." Og så: "Vil de af jer som ved I har mudder i panden kommer hen til mig." Hvis ingen kommer hen til faren, vil han gentage beskeden (igen og igen). Vil nogen af børnene nogensinde komme hen til faren?*

Antag netop m af børnene har mudder i panden ($m > 0$). Det kan vises at hvis børnene ræssonerer perfekt vil der ikke ske noget de første $m - 1$ gange faren giver beskeden, men den m'te gang vil netop de m mudrede børn træde frem!

Iteration af viden

Det ses at det kræves at børnenes ræssonement ikke kun involverer viden omkring den "ydre verden", men også viden om hvad de andre børn ved.

Vi kan altså ikke modellere viden realistisk uden at tillade viden om viden som f.eks. "jeg ved at Anders ved at jeg har højst 2 visdomstænder".

Viden om viden kaldes **metaviden**. Og når vi kan have viden om viden, kan vi naturligvis også have viden om viden om viden, osv. Viden kan itereres og danner en kæde af niveauer.

Formelt kan vi opnå dette ved at repræsentere viden ved en logisk operator som kan anvendes på vilkårlige udsagn, også udsagn der selv involverer viden:

"Jeg ved at to plus to er fire"	$\curvearrowright$	$K_{\mathsf{jeg}}(2+2=4)$
"Jeg ved at Anders ved at to plus to er fire"	$\curvearrowright$	$K_{\mathsf{jeg}}K_{\mathsf{anders}}(2+2=4)$
"Jeg ved at Anders ved at jeg ved at to plus to er fire"	$\curvearrowright$	$K_{\mathsf{jeg}}K_{\mathsf{anders}}K_{\mathsf{jeg}}(2+2=4)$

Theory of Mind

Iteration af viden som i $K_{\text{jeg}} K_{\text{anders}} K_{\text{jeg}}(2+2=4)$ kan synes lidt "søgt" og forekommer at være noget vi sjældent bruger i praksis. Men de mudrede børn kan ikke klare sig uden. I tilfældet med 3 børn hvoraf nummer 0 og 1 er mudrede, vil barn nummer 0 have behov for at kun ræssonere sig frem til følgende:

$$K_0(\neg \text{mudret}(0) \wedge K_1(\text{mudret}(0) \vee \text{mudret}(1) \vee \text{mudret}(2)) \rightarrow K_1(\text{mudret}(1))).$$

Og jo flere mudrede børn, jo dybere iterationer af viden er nødvendigt.

Et andet eksempel optræder i forbindelse med **Theory of Mind (ToM)** (Premack og Woodruff, 1978). Theory of Mind betegner vores evne til at tillægge andre væsener mentale tilstande, viden og overbevisninger. At have en Theory of Mind er naturligvis essentielt for succesfuld social omgang og sproglig kommunikation.

Eksempel. "Kunne du fortælle mig hvor jeg finder et apotek?"

Sally-Ann test

Eksperimenter viser at normale børn danner en Theory of Mind omkring 4 års-alderen (Wimmer and Perner, 1983). Autistiske børn har til gengæld ofte en defekt Theory of Mind (Baron-Cohen *et al.*, 1985).

En af de mest benyttede eksperimenter for at teste Theory of Mind er den såkaldte *Sally-Ann test*. Man fremfører med dukker følgende scenarie for forsøgspersonen:

> *Sally og Ann leger med en bold. Sally putter den i en kurv, og forlader derefter rummet. Imens flytter Ann bolden over i en anden beholder, f.eks. en kasse. Herefter kommer Sally tilbage.*

Man spørger nu forsøgspersonen: Hvor vil Sally lede efter bolden? Hvis man siger "i kurven" har man "bestået", ellers har man "dumpet".

Sally-Ann test

Pointen i Sally-Ann testen er at forsøgspersonen skal være i stand til at forstå at Sallys overbevisninger ikke stemmer overens med Anns og forsøgspersonens egen viden om verdens faktiske tilstand. Forsøgspersonen skal kunne ræssonere sig frem til følgende:

$$K_{\mathsf{jeg}}(\neg \mathsf{i\_kurv}(\mathsf{bold}) \wedge B_{\mathsf{Sally}} \mathsf{i\_kurv}(\mathsf{bold})).$$

Hvis vi ønsker at lave intelligente computer-systemer (robotter o.lign.) som kan matche 4-årige er vi altså som absolut minimum nødt til at have systemer som kan repræsentere og ræssonere med udsagn/modeller involverende itereret viden som ovenfor.

Syntaks for epistemisk modallogik

Lad os gå lidt i dybden med de logiske aspekter af hvordan viden skal formaliseres. Som sagt er det naturligt at se på viden og tro som logiske operatorer.

Et sprog med de sædvanlige udsagnslogiske operatorer udvidet med operatorer for viden og tro for hver mulig agent (person) kan se således ud:

$$\phi, \psi ::= p \mid \neg\phi \mid \phi \wedge \psi \mid B_a\phi \mid K_a\psi,$$

hvor p tilhører en predefineret mængde af propositionssymboler og a tilhører en predefineret mængde af agenter.

Syntaktisk tilgang til viden

Betragt igen den foreslåede syntaks:

$$\phi, \psi ::= p \mid \neg\phi \mid \phi \wedge \psi \mid B_a\phi \mid K_a\psi,$$

Det ligner unægteligt definitionen af syntaksen for modallogik.

Men det behøver ikke være en modallogik. Udtrykket $B_a\phi$ kan eksempelvis blot være en forkortet skrivemåde for $(B_a\phi)^\star$, hvor $\cdot^\star$ betegner en oversættelse ind i prædikatlogik:

$$
\begin{aligned}
&(B_a\phi)^\star \overset{df}{=} B(a, \ulcorner \phi^\star \urcorner) \\
&(K_a\phi)^\star \overset{df}{=} K(a, \ulcorner \phi^\star \urcorner) \\
&\psi^\star \overset{df}{=} \psi, \text{ for alle andre typer af formler } \psi.
\end{aligned}
$$

Her er B og K binære relationssymboler i prædikatlogik, og $\ulcorner \cdot \urcorner$ betegner en afbildning af formler ind i termer (en **Gödelkodning**).

Benyttes oversættelsen ind i prædikatlogik fås den såkaldt **syntaktiske tilgang** til viden og overbevisning.

Semantisk tilgang til viden

Betragt igen den foreslåede syntaks:

$$\phi, \psi ::= p \mid \neg\phi \mid \phi \wedge \psi \mid B_a\phi \mid K_a\psi,$$

Opfattes B_a og K_a som (modal-)operatorer fås den såkaldte **semantiske tilgang** til viden og overbevisning. Det er den vi vil forfølge her.

Vi har nu styr på en **syntaks** (hvilke sætninger der kan dannes), men har endnu ikke besluttet os for en **semantik** (fortolkningen af sætningerne).

Betragt eksemplet med 3 mudrede børn hvori 0 og 1 har mudder i panden. I starten vil 0 opfatte netop følgende to forskellige omstændigheder som mulige: "0 og 1 har mudder i panden, 2 er ren", "1 har mudder i panden, 0 og 2 er rene". Vi kan altså repræsentere 0's syn på verden som mængden af disse to mulige tilstande (de alternativer som agenten opfatter som mulige).

Denne idé antyder at vi uden videre kan benytte en almindelig **muligverdens-semantik** à la Kripke til at give betydning til vores epistemiske sætninger.

Muligverdens-semantik for epistemisk logik

En almindelig **muligverdens-semantik** for epistemisk logik ser ud som følger.

En **model** er en tuppel $(W, (R_x)_{x \text{ er modaloperator}}, V)$, hvor W kaldes mængden af **mulige verdener** (eller **tilstande**), alle R_x er binære **tilgængelighedsrelationer** på W og V afbilder propositionssymboler ind i delmængder af W (de verdener hvori propositionen er sand).

Givet en model $M = (W, R, V)$ defineres opfyldelighedsrelationen $M, w \models \phi$ mellem en model M, en verden w og en formel ϕ på følgende måde:

$$
\begin{aligned}
M, w \models p &\overset{df}{\Leftrightarrow} w \in V(p) \\
M, w \models \neg\phi &\overset{df}{\Leftrightarrow} \text{ikke } M, w \models \phi \\
M, w \models \phi \wedge \psi &\overset{df}{\Leftrightarrow} M, w \models \phi \text{ og } M, w \models \psi \\
M, w \models B_a\phi &\overset{df}{\Leftrightarrow} \text{for alle } v \in W \text{ med } (w, v) \in R_{B_a} \text{ gælder } M, v \models \phi \\
M, w \models K_a\phi &\overset{df}{\Leftrightarrow} \text{for alle } v \in W \text{ med } (w, v) \in R_{K_a} \text{ gælder } M, v \models \phi
\end{aligned}
$$

Gyldighed

Betragt igen definitionen af opfyldelighedsrelationen:

$$
\begin{array}{rcl}
M, w \models p & \overset{df}{\Leftrightarrow} & w \in V(p) \\
M, w \models \neg\phi & \overset{df}{\Leftrightarrow} & \text{ikke } M, w \models \phi \\
M, w \models \phi \wedge \psi & \overset{df}{\Leftrightarrow} & M, w \models \phi \text{ og } M, w \models \psi \\
M, w \models B_a\phi & \overset{df}{\Leftrightarrow} & \text{for alle } v \in W \text{ med } (w, v) \in R_{B_a} \text{ gælder } M, v \models \phi \\
M, w \models K_a\phi & \overset{df}{\Leftrightarrow} & \text{for alle } v \in W \text{ med } (w, v) \in R_{K_a} \text{ gælder } M, v \models \phi
\end{array}
$$

Vi siger at en formel ϕ er **gyldig** hvis $M, w \models \phi$ for alle modeller $M = (W, R, V)$ og verdener $w \in W$.

Eksempler på gyldige formler:

$$
\begin{array}{l}
K_a(\phi \wedge \psi) \rightarrow K_a\phi \\
K_b(K_a(\phi \wedge \psi) \rightarrow K_a\phi) \\
K_a\phi \rightarrow \neg K_a(\neg\phi) \\
K_a(\phi \rightarrow \psi) \wedge K_a(\phi) \rightarrow K_a(\psi).
\end{array}
$$

Kritik af muligverdens-semantik for epistemisk logik

Ideen om en muligverdens-semantik for epistemisk logik går tilbage til Hintikka, 1962. Denne semantik er i stadig i dag den eneste alment accepterede, selvom den dog har en række svagheder:

1. **Logisk omnisciens**. Det følger af den valgte semantik at hvis en agent ved både ϕ og $\phi \to \psi$ så ved han også ψ (se forgående slide). Det synes ikke realistisk, men er dog acceptabelt for idealiseret viden som ikke tager højde for ressource-begrænsninger.
2. Fravær af viden repræsenteres ved tilstedeværelsen af alle de mulige verdenstilstande som den manglende viden muliggør. Derfor gælder at "jo mindre man ved, desto større er ens model".

Litteraturen har primært kredset sig om den første svaghed og givet forskellige mulige løsninger på den (se f.eks. Fagin *et al.*, 1995).

Den anden svaghed synes dog måske endnu mere problematisk, i hvert fald hvis semantikken skal forestille sig at have en relation til vores kognitive repræsentation.

Ræssonering i epistemisk logik

Selvom muligverdens-semantikken ikke er uproblematisk i en epistemisk ramme er det dog alligevel den vi vil vælge her (i mangel af bedre).

Hvis man er et mudret barn eller en forsøgsperson i en Sally-Ann test har man brug for at **ræssonere** omkring formler i epistemisk logik for at kunne handle som tilsigtet.

Formelt set drejer det sig om at man ønsker at ræssonere sig frem til om en epistemisk formel ψ er sand eller ej, givet den baggrundsviden ϕ man har. Man ønsker altså at være i stand til at vurdere **gyldigheden** af epistemiske udsagn på formen $\phi \to \psi$.

Eksempel med ræssonering

Betragt barn nummer 0 i de mudrede børn efter faren er kommet med sin første besked og ingen træder frem. Da inkluderer 0's baggrundsviden følgende:

$$\begin{aligned}
&\text{mudder}(1) \wedge \neg\text{mudder}(2) \\
&K_i(\text{mudder}(0) \vee \text{mudder}(1) \vee \text{mudder}(2)), i = 1, 2 \\
&\text{mudder}(i) \leftrightarrow K_j(\text{mudder}(i)), i \neq j \\
&\neg\text{mudder}(i) \leftrightarrow K_j(\neg\text{mudder}(i)), i \neq j \\
&\neg K_1(\text{mudder}(1))
\end{aligned}$$

Kald konjunktionen af ovenstående formler for ϕ. Det kan vises at formlen $\phi \rightarrow \text{mudder}(0)$ er gyldig, og det er netop denne formels gyldighed som 0 skal ræssonere sig frem til for at kunne agere korrekt (træde frem) efter farens anden besked.

Men hvordan skal barnet foretage dette ræssonement? Med andre ord, hvordan bestemmes om en formel $\phi \rightarrow \psi$ i epistemisk logik er gyldig eller ej?

Bevisteori

Hvis vi forsøger at bestemme gyldigheden af en formel ϕ direkte udfra den definerede muligverdens-semantik har vi det problem at vi skal checke **alle** verdener i **alle** modeller. Det giver os kun en uendelig procedure som aldrig vil terminere medmindre vi tilfældigvis støder på en verden og en model hvori ϕ er falsk.

Vi har altså brug for noget bedre, og det er her at **bevisteori** kommer ind i billedet. I bevisteori studeres kalkuler/procedurer til at afgøre gyldigheden af formler. Der findes en lang forskellige typer af kalkuler og procedurer, hvoraf de vigtigste er:

- Hilbert-systemer.
- Naturlig deduktion.
- Sekvent-kalkuler.
- Tableau-kalkuler.
- Resolutions-systemer.

Mere om bevisteori

Betragt igen de forskellige kalkuler/procedurer til at afgøre gyldighed:

- Hilbert-systemer.
- Naturlig deduktion.
- Sekvent-kalkuler.
- Tableau-kalkuler.
- Resolutions-systemer.

For epistemisk logik (og modallogik i almindelighed) har især tableau-kalkuler og resolutions-systemer vist sig at kunne implementeres forholdsvist effektivt på computer ("forholdsvist" fordi man uafhængigt af sit program altid beviseligt vil kunne finde delklasser af formler, hvor den tid det tager at afgøre deres gyldighed vokser eksponentielt med formlens længde).

I de følgende ser vi på tableau-kalkuler som et eksempel.

Tableau-kalkuler for epistemisk logik

Antag at vi ønsker at undersøge gyldigheden af en formel ϕ ved hjælp af en tableau-procedure.

Ideen i tableau-proceduren minder på mange måder om den naive idé med blot at checke ϕ i alle modeller og verdener. Men man kan naturligvis aldrig blive færdig med at checke **alle** modeller og **alle** verdener, så hvordan skal vi nogensinde kunne vise eller afvise ϕ's gyldighed?

Tricket er at vi i stedet kan forsøge at lede efter blot **én** model og **én** verden hvori ϕ **ikke** er sand. Hvis det lykkes har vi vist at ϕ **ikke** er gyldig. Men hvad hvis det ikke lykkes?

Pointen er at vi systematisk kan udvikle alle de modeller og verdener hvori ϕ har muligheden for at være falsk. Hvis denne systematiske udvikling fejler kan vi så konkludere at der ikke eksisterer en sådan model og verden—og dermed at ϕ må være gyldig.

Det er ikke på forhånd klart at vi får en endelig procedure på denne måde. Men det kan vises at proceduren under passende restriktioner faktisk bliver endelig.

Tableau-kalkuler for epistemisk logik

Tableau-procedurer fungere således ved systematisk at lede efter en model af en given formel. Man opbygger trinvist sine model-kandidater udfra de betingelser som formlen dikterer.

Eksempel. Nedenfor er angivet et tableau for ræssonementet som leder barn nummer 0 frem til at han har mudder i panden. 0's baggrundsviden er angivet med grønt. Negationen af den ønskede konklusion er angivet med rødt. De i tableauet udledte formler er angivet med sort.

$$\begin{array}{ll}
m(1) \land \neg m(2) & \\
K_i(m(0) \lor m(1) \lor m(2)) & \\
m(i) \leftrightarrow K_j(m(i))), i \neq j & \\
\neg m(i) \leftrightarrow K_j(\neg m(i)), i \neq j & \\
\neg K_1(m(1)) & \\
\neg m(0) & \\
K_1(\neg m(0)) & 5 \\
\neg m(2) & 8 \\
K_1(\neg m(2)) & 9
\end{array}
\xrightarrow[1]{1}
\begin{array}{ll}
\neg m(1) & 2 \\
m(0) \lor m(1) \lor m(2) & 3 \\
m(0) \lor m(2) & 4 \\
\neg m(0) & 6 \\
\left.\begin{array}{ll} m(2) & 7 \\ \neg m(2) & 10 \end{array}\right\} & \text{modstrid!}
\end{array}$$

Epistemisk logik, naturligt sprog og spil

Luciana Benotti har i sit netop afsluttede ph.d.-projekt studeret et tekstbaseret adventure game som kommunikerer via naturligt sprog.

For at kunne give mening til og for at kunne udføre kommandoer som "åben døren", "lås døren op" og "lås døren op med nøglen" viser det sig at der kræves epistemisk repræsentation og ræssonementer.

I Benottis spil benyttes en tableau-baseret inferens-maskine til at håndtere de epistemiske ræssonementer.

Konklusioner

- Itereret viden er afgørende for korrekt at foretage en række ræssonementer såsom i de mudrede børn og Sally-Ann tests. Men også i en lang række mere realistiske og almindeligt forekommende scenarier involverende social interaktion og sproglig kommunikation.
- For at kunne lave effektive kunstig intelligens-systemer er det derfor nødvendigt at kunne håndtere epistemisk repræsentation og ræssonering. Dette kan gøres ved hjælp af epistemisk logik og dets bevisteori.
- Feltet epistemisk logik har udviklet sig meget kraftigt de sidste 15 år, men der er stadig mange løse ender og givetvis langt igen inden vi har kunstig intelligens-systemer som effektivt kan lave epistemiske ræssonementer udenfor meget afgrænsede og foruddefinerede domæner.