

SICS/T-89/8906

Två system för default-resonemang

En studie av Kurt Konoliges artikel:

On the Relation between Default and Autoepistemic Logic

**av
Per Kreuger**

Två system för default-resonemang

En studie av Kurt Konoliges artikel:
On the Relation between Default and Autoepistemic Logic

av
Per Kreuger

Abstract

Denna artikel presenterar, sammanfattar och granskar kritiskt en teori som läggs fram av Kurt Konolige i hans artikel "On the Relation between Default and Autoepistemic Logic" [Ko 87]. Detta material presenterades ursprungligen som examination på doktorandkursen "Logik for AI" som gavs under 1988 av Rune Gustavsson på SICS.

Konoliges artikel tar upp förhållandet mellan två typer av logiker som använts för att föra default-resonemang, en typ av resonemang under ofullständig kunskap. De två system han tar upp är dels Reiter's Default-logik [Re 80], dels Moore's Autoepistemiska logik [Mo 85]. Det huvudsakliga resultatet i Konolige's artikel är att dessa två system kan visas vara ekvivalenta i en viss mening.

1. Inledning

1.1. Default-resonemang

Default-resonemang bygger på konstaterandet att människor drar slutsatser inte bara utifrån vad de vet, utan också ifrån vad som normalt är fallet, d.v.s på antaganden de gör om världen. Detta är ett exempel på s.k. *ickemonotona* resonemang. De slutsatser människor drar på detta sätt måste återkallas om de senare får veta att något antagande på vilket de byggt sitt resonemang inte längre är giltigt.

1.1.1. Default-logik

Default-logik är ett försök till formalisering av default-resonemang. En default-teori $\langle W, D \rangle$ består av en agents kunskapsbas (W), och en uppsättning default-regler (D). Kunskapsbasen W består av en mängd formler ur ett första ordningens språk. Default-reglerna har den allmänna formen:

$$\frac{\alpha : M \beta}{\omega},$$

där regeln skall förstås som att vi sluter oss till ω med ett default-resonemang om α kan visas från KB, och β är konsistent med KB. D.v.s. ω följer vanligtvis ur α om vi inte känner till någonting (β) som skulle motsäga det.

default-reglerna opererar på en metateoretisk nivå, och kan inte sägas vara en del av det logiska språket. De kan ses som transformationer på kunskapsbaser, så att applikationen av en default-regel i själva verket avbildar W på en ny kunskapsbas W' .

Kan man formulera en (kompositionell) semantik för ett sådant språk?

Ett alternativ är att försöka beskriva default-reglerna i språket självt. Detta kräver att vi kan referera till kunskapsbasen som helhet med någon typ av konstruktion i vårt språk. En teori som innehåller en sådan konstruktion kallar vi en indexikal teori.

1.1.2. Autoepistemisk Logik

Ett försök som gjorts i den riktningen är s.k. Autoepistemisk logik som först demonstrerades av Robert Moore [Mo 85]. Autoepistemisk logik använder en modaloperator L för att referera till en agents mängd av trossatser. Satsen $L\alpha$ skall tolkas som att α ingår i agentens trosmängd. Default-regler kan formuleras i AE logik, t.ex. på följande sätt. Det default-logiska schemat ovan ersätts med någonting i stil med

$$\alpha \wedge \neg L \neg \beta \rightarrow \omega.$$

D.v.s. om α gäller, och vi inte vet (tror) att β är falskt så antar vi ω .

En stor fördel med AE logik är att Moore definierar en kompositionell semantik för språket, något som aldrig gjorts för default-logik.

1.1.3. Relationen mellan dessa

Vad är då relationen mellan dessa två språk? Kan allting som kan uttryckas i det ena också uttryckas även i det andra. Då språken är såpass olika är detta inte på något sätt uppenbart. Vi nämnde ett sätt att uttrycka vissa default regler i AE logik. Men gäller detta allmänt? Det visar sig att default-logiken faktiskt kan ersättas med AE-logik i alla fall. Med vissa inskränkningar kan faktiskt också satser i AE-logiken uttryckas i default-logik, vilket kanske är mer förvånande.

2. Autoepistemisk logik

Som redan nämnts är Autoepistemisk logik ett språk som tillåter formulering av indexikala teorier. Detta sker med hjälp av modal operatorm L . Den avsedda tolkningen av $L\phi$ är att agenten tror ϕ . Satser i språket kan uttrycka relationer mellan agentens trossatser och fakta i världen. Vi skall strax titta lite närmare på detta språk, men vi skall först notera två saker. Språket tillåter inte kvantifiering av variabler in i ett modalt kontext. Detta innebär att t.ex. satser av typen $\exists xL(P(x))$ explicit utesluts. Vi kan m.a.o t.ex. inte påstå att agenten tror någonting, utan att specifikt säga vad det är. Å andra sidan är det just detta som gör det möjligt att formulera en kompositionell semantik för språket. I och med denna inskränkning behåller vi den egenskap som på engelska brukar kallas "referential transparency". Det andra vi bör notera är att språket bara tillåter en agent. Det är möjligt att de skulle gå att generalisera det ramverk som här ges av Konolige till att hantera flera agenter, men det är inte trivialt hur det i så fall skulle ske.

2.1. Syntax

Låt \mathcal{L}_0 vara ett första ordningens språk med funktioner och en särskild symbol \perp , som alltid är falsk. Utöka språket med den enställiga modaloperatorm L . Kalla detta utvidgade språk \mathcal{L} . \mathcal{L} definieras rekursivt av:

1. Bildningsreglerna för \mathcal{L}_0 .
2. Om ϕ är en sluten utsaga i \mathcal{L} så är $L\phi$ det också.

En term $L\phi$ kallas en modal atom i språket \mathcal{L} . Detta är rimligt eftersom ϕ i $L\phi$ inte får innehålla några fria variabler.

2.2. Tolkningar

Tolkningen av satser i språket ges av en första ordningens värdering, där icke modala atomer tolkas på vanligt sätt, d.v.s. av en struktur $I = \langle \mathcal{U}, \varphi, \mathcal{R} \rangle$, där \mathcal{U} är ett universum (en mängd individer), φ en avbildning av termer till element i \mathcal{U} och \mathcal{R} en mängd relationer över \mathcal{U} . Modala atomer tolkas av en mängd Γ sådan att en modal atom:

$$L\phi \text{ är sann om } \phi \in \Gamma.$$

Vi skriver $\models_{I,\Gamma} \phi$ om ϕ är sann under $\langle I, \Gamma \rangle$, och $\models_{\Gamma} \phi$ om ϕ är sann för alla tillordningar av värden till ickemodala atomer I , och värderingen av modala atomer ges av Γ . Den intuitiva tolkningen av en sådan Γ är mängden av satser agenten anser (tror) vara sanna.

2.3. Autoepistemisk utvidgningar

Antag nu att vi vill modellera en introspektiv agent! Givet en mängd av trossatser och en konsekvensrelation, vilken teori kan en sådan agent då sägas omfatta? Om vi kallar den ursprungliga mängden satser för A skulle en rimlig kandidat till den teori T vi är ute efter vara:

$$T = \{ \varphi \mid A \models_T \varphi \}$$

Vi kallar T en autoepistemisk (AE-) utvidgning till A . Problemet med denna definition är att den är cirkulär. Teorin består alltså av alla satser som följer ur A givet den konsekvensrelation som genereras av T . Vi kan betrakta denna "cirkulära" definition av en trosmängd som en fixpunktsekvation i T .

Denna fixpunkts ekvation definierar T 'n som är sunda och fullständiga i förhållande till A i följande bemärkelse:

T är en sund utvidgning av A omm varje värdering $\langle J, T \rangle$ som är en modell till A också är en modell till T .

T är en semantiskt fullständig utvidgning av A omm T innehåller varje sats som är sann i varje AE-modell till T .

Moore kallar en utvidgning T av en mängd satser A stabil om den uppfyller ovanstående två krav. Alla stabila utvidgningar är AE-utvidgningar och omvänt.

Om A inte innehåller några modala atomer så har den exakt en AE-utvidgning

För att visa relationen mellan AE-logik och Default-logik skall vi nu analysera dess AE-utvidgningar lite närmare. Det visar sig att vi kan ge en alternativ karaktärisering av T på följande sätt.

$$\text{Låt } LT = \{ L\varphi \mid \varphi \in T \} \text{ och } \neg L\overline{T} = \{ \neg L\varphi \mid \varphi \notin T \},$$

$T \setminus (LT \cup \neg L\overline{T})$ innehåller då inga modala atomer. Vi kan också se att

$$LT \cup \neg L\overline{T} \models \varphi \text{ omm } \models_T \varphi$$

I och med detta kan vi flytta den självrefererande delen av definitionen från konsekvensrelationen till antagandena, och använda vanlig logisk konsekvens som konsekvensrelation.

$$T = \{ \varphi \mid A \cup LT \cup \neg L\overline{T} \models \varphi \}.$$

Denna definition karaktäriserar samma mängd T som den ovan.

2.4. Stabila mängder och grundade utvidgningar

En stabil mängd är (efter Moore) en mängd av formler ur ett första ordningens språk med en modal operator L sådan att:

1. $\phi \in \Gamma$ om $\phi \models \Gamma$ Slutet under logisk konsekvens
2. $L\phi \in \Gamma$ om $\phi \in \Gamma$ Positiv introspektion
3. $\neg L\phi \in \Gamma$ om $\phi \notin \Gamma$ Negativ introspektion

Frågan är ju nu om dessa mängder är intressanta, om de har egenskaper som tro har. Låt oss lämna denna fråga öppen t.s.v., och se lite närmare på dessa mängder.

Konolige visar att AE-utvidgningar och stabila mängder är relaterade på följande sätt:

1. Varje AE-utvidgning av A är en stabil mängd som innehåller A .
2. Varje stabil mängd är en AE-utvidgning av sina icke-modala satser
3. Om W är en mängd icke-modala formler så existerar det en unik stabil mängd Γ så att $\Gamma_0 = W$, där Γ_0 är de icke-modala formlerna ur Γ . W kallas kärnan till Γ .

Allt detta ger oss en alternativ karaktärisering av AE-utvidgningar, nämligen:

Låt \models_{SS} vara den konsekvensrelation som fås genom att inskränka \models_{Γ} till sådana Γ som är stabila. $\models_{SS} \phi$ betyder alltså att ϕ är sann under alla tolkningar Γ av de modala atomerna sådan att Γ är en stabil mängd.

Eftersom varje stabil mängd genereras av sina icke-modala formler får vi då:

$$T = \{\phi \mid A \cup LT_0 \cup \neg \overline{LT_0} \models_{SS} \phi\},$$

där T_0 är mängden av ickemodala formler i T , och $\overline{T_0}$ är mängden av ickemodala formler $\neg \phi$ i T .

Har vi nu eliminerat cirkuläriteten i vår definition? Det kan synas så, men vi har fortfarande en relation mellan T_0 , $\neg T_0$ och T .

Vi kan nu inskränka vår definition av AE-utvidgningar för att få en mer intuitivt tilltalande modell för introspektiva agenter.

$$T = \{\phi \mid A \cup LA \cup \neg \overline{LT_0} \models_{SS} \phi\},$$

är en inskränkning av karaktäriseringen ovan, som utesluter att agenten använder formler av typen $LP \rightarrow P$, för att sluta sig till P . Detta skulle motsvara att en agent säger sig tro någonting med motiveringen att han tror att han tror det.

Konolige kallar detta krav på AE-utvidgningen att utvidgningen är (moderat¹) grun-

1. Det finns en svagt grundad utvidgning också, som alla AE-utvidgningar uppfyller, samt en starkt grundad utvidgning, som är ett tekniskt krav för att få igenom ekvivalensen mellan de båda systemen. Jag går inte igenom någondera i detalj.

dad i A.

2.5. Bevisteori

När vi nu betraktar konsekvensrelationen \models_{SS} ser vi att den överensstämmer exakt med den (epistemiska, doxastiska) modal logik som kallas $K45$, och som konstruerats just för att modellera tro snarare än vetande. Jag går inte in närmare på hur detta resultat bevisas, men det bygger på kopplingen mellan stabila mängder, och möjliga världar i Kripke-semantiken för $S5$. Givet detta resultat kan vi börja konstruera en bevisteori för AE-logik. Syftet med detta är inte främst att få fram en (semi-)automatisk bevismetod, även om detta naturligtvis också är önskvärt, utan för att underbygga det huvudsakliga resultatet i artikeln.

Jag ger här (för fullständighetens skull) det logiska systemet för $K45$.

$L(\phi \rightarrow \psi) \rightarrow (L\phi \rightarrow L\psi)$	Distrb.
$L\phi \rightarrow LL\phi$	4 (Positive intösp.)
$\neg L\phi \rightarrow L\neg L\phi$	5 (Negativ intösp.)
$\frac{\phi \quad \phi \rightarrow \psi}{\psi}$	Modus Ponens
$\frac{\phi}{L\phi}$	Necessitation

En bevisteoretisk framställning av AE-logiken skulle då kunna var i termer av bevisteorin för $K45$:

$$T = \{\phi \mid A \cup LT_0 \cup \neg \overline{LT}_0 \vdash_{K45} \phi\},$$

eller för det (moderat) grundade fallet,

$$T = \{\phi \mid A \cup LA \cup \neg \overline{LT}_0 \vdash_{K45} \phi\},.$$

2.6. Normalform

Givet denna likhet med $K45$, kan vi utnyttja egenskaper hos $K45$ -system för att definiera en normalform för uttryck i AE-logik.

Vi kan med hjälp av dessa egenskaper etablera följande viktiga teorem:

För varje mängd av satser A i vårt språk \mathcal{L} finns en $K45$ -ekvivalent mängd av satser på formen:

$$\neg L\alpha \vee L\beta_1 \vee \dots \vee L\beta_n \vee \omega,,$$

där $K45$ -ekvivalens mellan två formler ϕ och ψ definieras som $\vdash_{K45} \phi \leftrightarrow \psi$

$\vdash_{K45} \psi$ och där α , β_i samt ω är icke-modala formler.

3. Default-logik

Som vi redan nämnt består en Default-teori $\langle W, D \rangle$ av en mängd formler i ett första ordningens språk; W och en mängd default-regler; D .

I sin mest generella formulering har en default-regel har den allmänna formen:

$$\frac{\alpha : M\beta_1, M\beta_2, \dots, M\beta_n}{\omega}$$

En default-regel är uppfylld av en mängd formler Γ om antingen

1. $\alpha \notin \Gamma$ eller något $\neg\beta \in \Gamma$ eller
2. $\omega \in \Gamma$

En default-utvidgning av en default-teori $\langle W, D \rangle$ är den minsta mängd Γ som innehåller W , är sluten under första ordningens konsekvens och som uppfyller alla default-regler i D .

Man kan betrakta en default-utvidgning som en fixpunkt till operatoren $\lambda V. \Gamma(V)$. Låt V vara en godtycklig mängd första ordningens formler och definiera operatoren Γ :

- D1. $W \subseteq \Gamma(V)$
- D2. För alla φ gäller att om $\Gamma(V) \models \varphi$ så $\varphi \in \Gamma(V)$

Default-utvidgningar är fixpunkter till operatoren $\lambda V. \Gamma(V)$, d.v.s. någon mängd E sådan att $\Gamma(E) = E$.²

2. Jmf. parametern V i $\Gamma(V)$ med mängden $\neg\overline{LT_0}$!

4. Relation mellan dessa två

Så kommer vi då till det huvudsakliga resultatet i Konoliges artikel – Jämförelsen mellan dessa båda system. Om vi bara betraktar de två språken var och en för sig, och tar hänsyn till deras tänkta användning finns ingen uppenbar likhet.

AE-logiken är ett första ordningens språk utvidgat med en modal operator, medan default-logik innehåller en mängd härledningsregler (default-reglerna), som i själva verket är formel-scheman, eller andra ordningens formler. För en fix uppsättning defaults kan en default-logik betraktas som ett första ordningens system, men så snart man betraktar default-reglerna som en del av språket går man in i ett andra ordningens system.

Vi kan ändå ställa oss frågan om alla satser i det ena av dessa språk kan uttryckas helt i det andra. Det visar sig att default-logik kan kodas i AE-logik, och att AE-logik (nästan) kan uttryckas i default-logik. Detta sista resultat är det mest förvånande, och enligt min mening en sanning med viss modifikation.

För att klara översättningen från AE-logik till default-logik måste vi inskränka AE-utvidgningar till de som Konolige kallar starkt grundade utvidgningar. Detta är ett mycket tekniskt krav.

Detta krav fanns inte med i de första versionerna av Konoliges artikel. Han nöjde sig då med att inskränka sig till de moderat grundade utvidgningarna vilket är ett betydligt rimligare krav. De moderat grundade utvidgningarna motsvarar i någon mån vår intuition om vad en idealiserad agent tror givet en mängd grund-trossatser. Detta krav visade sig sedan inte räcka, och kravet på starkt grundade utvidgningar är en ad hoc lösning på detta.

Jag går inte närmare in på hur kravet på de starkt grundade utvidgningarna ser ut. Det räcker i detta sammanhang att nämna att det är ett krav på den syntaktiska formen på meningar i normalform ur språket \mathcal{L} . Detta krav formuleras endast med en referens till denna syntaktiska form, vilket gör att vi måste uttala oss om en helt ny domän förutom de mängder vi normalt rör oss med då vi definierar modellteori för AE-logik.

Min slutsats av detta är att default-logik mycket väl kan uttryckas i AE-logik, medan motsatsen bara nästan är fallet. Jag tycker inte att detta gör Konoliges resultat mindre intressanta. Han har grundligt utrett sambandet mellan dessa båda språk, och där med gett en fördjupad förståelse av dem båda.

Jag skall nu ge översättningarna åt båda hållen. Håll då bara i minnet att bevisen för att översättningen från AE-logik till default-logik är korrekta beror på att AE-utvidgningarna inskränks till sådana som är starkt grundade.

4.1. Översättningar

Definiera en transformation som givet en default-teori $\langle W, D \rangle$ ger en AE-teori T , som följer!

Låt alla formler i W ingå i T .

Låt varje default-regel i D översättas enligt följande schema, och låt den resulterande formeln ingå i T .

$$\frac{\alpha : M\beta_1, \dots, M\beta_n}{\omega} \triangleright (L\alpha \wedge \neg L\neg\beta_1 \wedge \dots \wedge \neg L\neg\beta_n) \rightarrow \omega$$

Låt T vara den minsta mängd formler ur \mathcal{L} som uppfyller ovanstående krav.

Vi har då följande resultat:

Låt A vara resultatet av en transformation som ovan av en default-teori Δ . En mängd E är en default-utvidgning av Δ om den också är kärnan i en starkt grundad AE-utvidgning av A .

En starkt grundad AE-utvidgning är alltid en moderat grundad AE-utvidgning, så det existerar också en sådan.

Om vi inskränker oss till de starkt grundade AE-utvidgningarna vet vi att kärnan i en sådan teori alltid kan skrivas på en form som motsvarar resultatet av översättningen av default-reglerna ovan. Det är nämligen precis detta som kravet på stark grundadhet garanterar. Väldigt många moderat grundade utvidgningar har också en kärna vars normalform motsvarar denna, så kravet är inte orimligt, bara ointuitivt.

Sammantaget ger det oss en procedur för att översätta en mängd formler i AE-logik, till en default-teori $\langle W, D \rangle$ så att kärnan E i en starkt grundad AE-utvidgning av A är identisk med en default-utvidgning av $\langle W, D \rangle$ och omvänt.

4.2. Semantik för default-logik

En konsekvens av detta resultat är givetvis också att default-logik kan ges en kompositionell semantik på samma sätt som AE-logik. Detta ökar vår förståelse av default-logiken, och placerar in den i ett sammanhang av många andra språk som kan ges en semantik på liknande sätt.

5. Sammanfattning och slutsatser

Konoliges artikel visar alltså att det finns intressanta samband mellan default-logik, och AE-logik över språket \mathcal{L} . Det är klart att AE-logiken har hela default-logikens uttryckskraft, och att även omvändningen gäller om man inskränker AE-logiken på ett lämpligt sätt.

Man bör kanske också komma ihåg de inskränkningar vi gjorde då vi definierade språket \mathcal{L} . Det är högst sannolikt att översätt från AE-logik till default-logik kollapsar om vi tillåter kvantifiering över variabler i ett modalt kontext. Gör vi detta förlorar vi också den vackra och regelbundna semantik den AE-logik vi studerat här har. Likaså är det inte uppenbart vad som händer om vi istället inför ett språk med en tvåställig modaloperator för att kunna tala om flera agenter tro. Det troliga är att översättningen även då bryter samman.

Artikeln innehåller också en analys av de båda systemen som ger en belysning inte bara av systemens relation, utan också av de båda språken vart och ett för sig. Den skiss av en bevis teori för AE-logik som Konolige ger låter oss också ana att det kanske inte är alldeles omöjligt att automatisera resonemang i AE-logik. I vart fall inte mycket svårare än i modallogiken *K45*.

Vad innebär det då på ett mer filosofiskt plan att default-resonemang kan uttryckas som resonemang om introspektiva agenter? Låt oss gå tillbaka till översättningsschemat och betrakta det en gång till!

$$\frac{\alpha : M\beta_1, \dots, M\beta_n}{\omega} \triangleright (L\alpha \wedge \neg L\neg\beta_1 \wedge \dots \wedge \neg L\neg\beta_n) \rightarrow \omega$$

Tag det klassiska exemplet med Koko!

$$\frac{bird(Koko) : M\neg(pingvin(Koko)), M\neg(struts(Koko))}{flyger(Koko)} \triangleright$$

$$(Lbird(Koko) \wedge \neg Lpingvin(Koko) \wedge \neg Lstruts(Koko) \rightarrow flyger(Koko))$$

Om Koko är en fågel och det är konsistent att anta att Koko inte är en pingvin och det är konsistent att anta att Koko inte är en struts så sluter vi oss till att Koko kan flyga...

och

Om vi tror att Koko är en fågel och vi inte tror att Koko är en pingvin och vi inte tror att Koko är en struts så kan vi sluta oss till att Koko kan flyga.

Den enda skillnaden tycks vara att vi när vi uttalar os om konsistens i det default-logiska fallet refererar till en mängd av satser, men inte till någon agent, medan i AE-fallet säger vi explicit att den mängd vi refererar till är just den som omfattas av en agent. Kanske är det i själva verket så att när vi tänker på default-resonemang försöker vi modellera en tänkt agents icke-monotona slutledningar, trots att detta aldrig görs explicit i teorin.

I så fall vore det kanske naturligare att alltid formulera default-resonemang i AE-teorier, där denna referens görs explicit. Å andra sidan ger de resultat artikeln presenterar ett berättigande till klassisk default-logik, i och med att den ges en semantik via AE-logiken.

Det skulle vara intressant att betrakta några av de paradoxer som lätt uppstår i default-logik, för att se hur deras översättning till AE-logik skulle se ut. Översättningen ger ju inte en entydig utvidgning, och i några enkla fall kvar står paradoxen därför tämligen oförändrad. Det är dock möjligt att den intuition som legat till grund för den paradoxala formuleringen i default-logik lättare kan uttryckas i AE-logik, och paradoxen därmed undvikas. Detta är ett uppslag för vidare forskning.

- Ko 87 Konolige, K., "On the Relation between Default and Autoepistemic Logic" in Ginsburg, M. L. (ed.), *Readings in Nonmonotonic Reasoning*. Los Altos: Morgan Kaufman 1987.
- Mo 85 Moore, R. C., "A Formal Theory of Knowledge and Action" in Hobbes, J. R. and Moore, R. C. (eds.), *Formal Theories of the Commonsense World*. Norwood, NJ: Ablex, 1985
- Re 80 Reiter, R., "A Logic for Default Reasoning", *Artificial Intelligence*, 13(1-2): 81 - 132, 1980.