Jaké vlastnosti mají znalosti

(v sémantice možných světů)

Jeden ze způsobů jak charakterizovat naši interpretaci znalostí je popsat formule, které jsou vždy platné (pravdivé.)

Je-li dána struktura

$$M = (S, \pi, K_1, \ldots, K_n)$$

(i) říkáme, že *formule* ***A*** *platí (validní) v* ***M*** **(*M*** |= ***A*** **)**, jestliže v každém stavu ***s*** je ***A*** pravdivá **,** ***tj.***(*M*, *s*) |= ***A*** **.**

(ii) říkáme, že *formule* ***A*** *je splnitelná v* *M* , jestliže **(*M*, *t*)** |= ***A*** v nějakém stavu ***t*** .

(iii) říkáme, že ***formule A je platná (validní)*** a píšeme |= ***A*** , je-li platná (validní) ve všech strukturách.

(iv) říkáme, že *formule* ***A*** *je splnitelná*, je-li splnitelná v nějaké struktuře.

Platí

formule ***A*** je platná (platná v M) právě když formule
¬A není splnitelná (není splnitelná v M).

Připomeňme, že stále předpokládáme, že relace K_i jsou ekvivalence.

Validních formulí je jistě velké množství (všechny výrokové tautologie, …)

Hledáme nějaký způsob, jak je charakterizovat pomocí syntaktických prostředků.

Existuje FORMÁLMÍ SYSTÉM, který to dokáže?

Při zavedeném pojetí znalostí platí, že agent zná všechny logické důsledky svých znalostí, neboť platí:

Ví-li agent tvrzení A a také ví, že A implikuje B, potom obě formule A a $A \rightarrow B$ jsou pravdivé ve všech světech, které agent považuje za možné,

tedy také B musí být pravdivé ve všech světech, které agent považuje za možné, takže musí vědět (znát) i B .

Odtud dostáváme $\models (K_i A \wedge K_i(A \rightarrow B)) \rightarrow K_i B$

Této formuli se říká **axiom distribuce** nebo také Kripkův axiom a označuje se jako **K**, protože dovoluje distribuovat operátor K_i přes implikaci.

Druhá vlastnost našeho pojetí znalosti zaručuje, že máme dostatečně silné a schopné agenty: předpokládáme totiž, že každý agent zná všechny formule platné (validní) v dané struktuře.

Je-li ***A*** pravdivá ve všech stavech (tj. možných světech) struktury ***M***, pak je pravdivá ve všech těch světech, které agent v daném světě považuje za možné. TEDY:

*Pro libovolnou strukturu **M** platí*

je-li $\boldsymbol{M} \models \boldsymbol{A}$, *pak* $\boldsymbol{M} \models \boldsymbol{K_i A}$

Tak získáváme ***Pravidlo generalizace znalostí :*** $\dfrac{\vdash A}{\vdash K_i A}$

Pozor! Pravidlo generalizace je něco jiného než implikace

$$A \rightarrow K_i A$$

která tvrdí „je-li *A* pravdivá, potom agent *i* to ví“: toto **NENÍ validní formule**! Agent nemusí vědět všechny věci, které jsou pravdivé.

{*Například při hře zablácených dětí může mít jedno z nich zablácené čelo, ale nemusí to vědět.*}

Agenti znají všechny validní formule. Jinými slovy znají formule, které jsou *nutně* pravdivé.

To však neplatí o formulích, které jsou (*řízením osudu*) pravdivé jen v nějakém z možných světů.

Agent nemusí znát všechna fakta, která jsou pravdivá! Ale když už agent něco ví, pak to platí:

$$\models K_i A \rightarrow A$$

Tato vlastnost se obvykle nazývá ***Axiom znalostí*** nebo ***Axiom pravdy*** a často se označuje **T**.

Odůvodnění axiomu: svět, ve kterém se agent nachází, je vždy takový, který pokládá za možný. Platí-li $K_i A$ v nějakém světě (M, s), pak A platí ve všech světech, které i považuje za možné, tedy i v (M, s).

{Filosofové tento axiom považují za hlavní odlišení mezi **znalostí** a **přesvědčením** (vírou, *belief*). Mohu mít nepravdivé přesvědčení, ale nemohu vědět něco, co není pravda!}

V případě, že chceme popisovat to, čemu agent věří, nikoliv to, co ví, nahradíme axiom pravdy

$$\models K_i A \rightarrow A$$

slabším požadavkem:

Axiom konzistence požaduje $\neg K_i \mathit{false}$, který se často se označuje jako **D**.

Další dvě vlastnosti se týkají přirozeného požadavku, aby agenti věděli o svých znalostech (pomocí introspekce).

Agenti vědí, co vědí a vědí, co nevědí.

$$\models K_i A \rightarrow K_i K_i A$$
$$\models \neg K_i A \rightarrow K_i \neg K_i A$$

První z těchto vlastností se jmenuje ***Axiom pozitivní introspekce*** (často označovaný **4**),

druhá ***Axiom negativní introspekce*** (často označovaný (často označovaný **5**).

Označme $\mathcal{M}_n(\Phi)$ množinu všech Kripkeho struktur nad množinou prvotních formulí $\mathbf{\Phi}$ pro $\boldsymbol{n}$ agentů takových, že na relace $\boldsymbol{K_i}$ nejsou kladeny žádné požadavky.

$\mathcal{M}_n^{rst}(\Phi)$ nechť je pak podmnožina $\mathcal{M}_n(\Phi)$, ve které každá relace přípustnosti splňuje vlastnosti ***rst***:

reflexivní

symetrická

transitivní.

Jedná se tedy o relace **ekvivalence**.

Věta.

Pro všechny struktury M pro n agentů s relacemi K_i, které jsou ekvivalencemi, a pro libovolné formule A, B platí

$$(i)\quad M \models (K_i A \wedge K_i(A \to B)) \to K_i B$$

$$(ii)\quad je-li \quad M \models A \quad potom \quad M \models K_i A$$

$$(iii)\quad M \models K_i A \to A$$

$$(iv)\quad M \models K_i A \to K_i K_i A$$

$$(v)\quad M \models \neg K_i A \to K_i \neg K_i A$$

Formální (axiomatický) systém K_n

Axiomy: **A1.** Všechny tautologie výrokové logiky

A2. $(K_i\alpha \wedge K_i(\alpha \to \beta)) \to K_i(\beta)$

Odvozovací pravidla:

R1. Z α formulí $\alpha \to \beta$ odvoďte β (modus ponens)

R2. Z formule α odvoď $K_i\,\alpha$ (pravidlo zobecnění znalosti)

V **K_n** lze dokázat například tvrzení $K_i(\alpha \wedge \beta) \to K_i\alpha$ (viz dále).

Značení **pro dokazatelnost:** $K_n \vdash K_i(\alpha \wedge \beta) \to K_i\alpha$

Věta:

Systém **K_n** představuje **korektní a úplný** syntaktický **popis** všech formulí validních v množině $\mathcal{M}_n(\Phi)$ Kripkeho struktur.

Axiomy modální logiky

1. **Výrokové tautologie**
2. **Distribuční axiom** (označovaný někdy jako **K**)
   $(K_i A \wedge K_i (A \rightarrow B)) \rightarrow K_i B$
3. **Axiom znalostí (Axiom pravdy)** (ozn. jako **T**) $K_i A \rightarrow A$ r
4. **Axiom pozitivní introspekce** (ozn. jako **4**) $K_i A \rightarrow K_i K_i A$ t
5. **Axiom negativní introspekce** (ozn. jako **5**) $\neg K_i A \rightarrow K_i \neg K_i A$ s +t
6. **Axiom konzistence** (ozn.jako **D**) $\neg K_i\, false$

Odvozovací pravidla:
R1. Z $\boldsymbol{\alpha}$ formulí $\boldsymbol{\alpha \rightarrow \beta}$ odvoďte $\boldsymbol{\beta}$ (modus ponens)
R2. Z formule $\boldsymbol{\alpha}$ odvoď $\boldsymbol{K_i\, \alpha}$ (pravidlo zobecnění znalosti)

$K_n \vdash K_i(\alpha \wedge \beta) \rightarrow K_i \alpha$:

Důkaz: [Jako zdůvodnění **musíme** uvést buď ***odvolání na axiom*** nebo ***informaci o použití odvoz. pravidla*** na předchozí řádky důkazu**]**

1. $(\alpha \wedge \beta) \rightarrow \alpha$ **[Výroková tautologie]**

2. $K_i((\alpha \wedge \beta) \rightarrow \alpha)$ [**R2:** 1, čili **R2** použito na formuli z řádky 1 **]**

3. $(K_i(\alpha \wedge \beta) \wedge K_i((\alpha \wedge \beta) \rightarrow \alpha)) \rightarrow K_i \alpha$ **[K]**

4. $((K_i(\alpha \wedge \beta) \wedge K_i((\alpha \wedge \beta) \rightarrow \alpha)) \rightarrow K_i \alpha)$

$\rightarrow (K_i((\alpha \wedge \beta) \rightarrow \alpha) \rightarrow (K_i(\alpha \wedge \beta) \rightarrow K_i \alpha))$

[Výr. tautologie $((p \wedge q) \rightarrow r) \rightarrow (q \rightarrow (p \rightarrow r))$]

5. $K_i((\alpha \wedge \beta) \rightarrow \alpha) \rightarrow (K_i(\alpha \wedge \beta) \rightarrow K_i \alpha)$ [**R1:** 3,4**]**

6. $K_i(\alpha \wedge \beta) \rightarrow K_i \alpha$ [**R1:** 2,5**]**

$\mathbf{K}_n \vdash K_i(\alpha \wedge \beta) \equiv K_i\alpha \wedge K_i\beta$

1. $\mathbf{K}_n \vdash K_i(\alpha \wedge \beta) \to K_i\alpha$ [viz předchozí strana]
2. $\mathbf{K}_n \vdash K_i(\alpha \wedge \beta) \to K_i\beta$ [důkaz obdobný tomu na předchozí straně]
3. $(K_i(\alpha \wedge \beta) \to K_i\alpha) \to ((K_i(\alpha \wedge \beta) \to K_i\beta) \to (K_i(\alpha \wedge \beta) \to (K_i\alpha \wedge K_i\beta)))$
   [$(\rho \to \varphi) \to ((\rho \to \psi) \to (\rho \to (\varphi \wedge \psi)))$ výroková tautologie]
4. $(K_i(\alpha \wedge \beta) \to K_i\beta) \to (K_i(\alpha \wedge \beta) \to (K_i\alpha \wedge K_i\beta))$ [MP: 3,1]
5. $K_i(\alpha \wedge \beta) \to (K_i\alpha \wedge K_i\beta)$ [MP: 4,2]

Tvrzení 1: $\mathbf{K}_n \vdash K_i(\alpha \wedge \beta) \equiv K_i\alpha \wedge K_i\beta$

Důkaz: kombinace důkazů na této a předchozí straně.

Věta 1: Systém K_n představuje korektní a úplný syntaktický popis všech formulí validních v množině $\mathcal{M}_n(\Phi)$ Kripkeho struktur (K_n je axiomatizací vzhledem k $\mathcal{M}_n(\Phi)$).

Věta 2:

Axiom T má tvar $K_iA \to A$. Systém $\mathrm{T}_n = (\mathrm{K}_n$ + axiom T) je axiomatizací vzhledem k $\mathcal{M}_n^{r}(\Phi)$.

Axiom 4 má tvar $K_iA \to K_iK_iA$. Systém $\mathrm{S4}_n = (\mathrm{T}_n$ + axiom 4) je axiomatizací vzhledem k $\mathcal{M}_n^{rt}(\Phi)$.

Axiom 5 má tvar $\neg K_iA \to K_i\neg K_iA$. Systém $\mathrm{S5}_n = (\mathrm{S4}_n$ + axiom 5) je axiomatizací vzhledem k $\mathcal{M}_n^{rts}(\Phi)$.

.

Následující vztahy lze dokázat (příklady viz dále):

- $(K_n + A6) \vdash \neg (K_i \alpha \wedge K_i \neg \alpha)$.
- $\neg K_i (\alpha \wedge K_i \neg \alpha) \equiv \neg K_i \alpha \vee \neg K_i \neg K_i \alpha$
- $K_i (\alpha \wedge \beta) \equiv K_i \alpha \wedge K_i \beta$.

Další příklad dokazatelného tvrzení: K_n , $(\varphi \to \psi) \vdash K_i \varphi \to K_i \psi$

1. $(\varphi \to \psi)$ [výchozí formule]
2. $K_i(\varphi \to \psi)$ [R2]
3. $K_i\varphi \to (K_i(\varphi \to \psi) \to K_i\psi)$ [K2]
4. $(K_i\varphi \to (K_i(\varphi \to \psi) \to K_i\psi)) \to (K_i(\varphi \to \psi) \to (K_i\varphi \to K_i\psi))$
   [Prop-T1: $(\varphi \to (\psi \to \tau)) \to (\psi \to (\varphi \to \tau))$]
5. $K_i(\varphi \to \psi) \to (K_i\varphi \to K_i\psi)$ [MP 4 a 3]
6. $(K_i\varphi \to K_i\psi)$ [MP 5 a 2]

Tvrzení 2: Jsou-li dvě výrokové φ , ψ formule ekvivalentní (tj. $(\varphi \to \psi) \wedge (\psi \to \varphi)$ je tautologie, značeno $\varphi \equiv \psi$), pak K_n |- $K_i \varphi \equiv K_i \psi$.

Důsledek tvrzení dokázaných na této a předchozí stránce.

Následující vztahy lze dokázat:

a) $\mathbf{K}_n \vdash K_i(\alpha \wedge \beta) \equiv K_i \alpha \wedge K_i \beta$

b) $(K_n + A6) \vdash \neg (K_i \alpha \wedge K_i \neg \alpha)$

c) $(K_n + A3) \vdash A6$

d) $\mathbf{K}_n \vdash K_i \neg (p \rightarrow K_i p) \equiv K_i (p \wedge \neg K_i p) \equiv (K_i p \wedge K_i(\neg K_i p))$

e) V $(K_n + A3)$ nemůže být dokazatelné $K_i \neg (p \rightarrow K_i p)$

Další příklady důkazu některých platných vztahů:

c1) K2, T(Axiom 3)$\vdash \neg K_i$ *false*

1. K_i *false* $\rightarrow$ *false* [A3]
2. $\neg$ *false* $\rightarrow (\neg K_i$ *false*) [ekv. úprava 1]
3. $\neg K_i$ *false* [použití výrokové tautologie na 2]

c2) K2, T$\vdash \neg K_i \alpha \vee \neg K_i \neg K_i \alpha$

c3) K2, T$\vdash \neg K_i (\alpha \wedge \neg K_i \alpha)$

1. $K_i \neg K_i \alpha \rightarrow \neg K_i \alpha$ (A3, axiom pravdy)
2. $\neg K_i \neg K_i \alpha \vee \neg K_i \alpha$ (výrok. Tautologie – přepis 1), *viz* a1
3. $\neg (K_i \neg K_i \alpha \wedge K_i \alpha)$ (výrok. Tautologie – přepis 2)
4. $\neg K_i (\neg K_i \alpha \wedge \alpha)$ (tvrzení o ekvivalenci a) pro formuli 3), *viz* a2

Následující vztahy lze dokázat (viz cvičení):

a) $\mathbf{K}_n \vdash K_i(\alpha \wedge \beta) \equiv K_i\,\alpha \wedge K_i\,\beta$

b) $(\mathrm{K}_n + \mathrm{A6}) \vdash \neg\,(K_i\,\alpha \wedge K_i \neg\,\alpha)$

c) $(\mathrm{K}_n + \mathrm{A3}) \vdash \mathrm{A6}$

d) $\mathbf{K}_n \vdash K_i \neg\,(p \rightarrow K_i\,p) \equiv K_i\,(p \wedge \neg\,K_i\,p) \equiv (K_i\,p \wedge K_i\,(\neg\,K_i\,p))$

e) V $(\mathrm{K}_n + \mathrm{A3})$ nemůže být dokazatelné $K_i \neg\,(p \rightarrow K_i\,p)$

$\boxed{E_G \quad C_G \quad D_G}$

G je podmnožinou {1, 2, . . . , n}, ***$E_G A$*** je **pravdivé**, právě když každý agent z ***G*** ví ***A***. Tedy

Axiom C1. $E_G A \Leftrightarrow \bigcap_{i \in G} K_i A$

Intuitivně, **společná** (všeobecná, *common*) **znalost** je to „*co je každému zcela jasné a ví to*", proto nepřekvapí, že společná znalost má všechny vlastnosti znalostí podobné **Distribučnímu axiomu**, **Axiomu znalostí**, **Axiomům pozitivní** a **negativní introspekce**. (Cvičení)

Pro všeobecnou znalost mezi skupinami agentů platí

$$Je-li \quad Q \subseteq G \quad potom \quad C_G A \rightarrow C_Q A$$

Cvičení. Dokažte, že následující formule jsou validní:

(i) $(C_G A \wedge C_G(A \to B)) \to C_G B$

(ii) $C_G A \to A$

(iii) $C_G A \to C_G C_G A$

(iv) $\neg C_G A \to C_G \neg C_G A$

Předpoklady o relacích K_i jsou stejné, jako při odvození odpovídajících axiomů.

Axiom o pevném bodu.

Nechť p je výrok „alespoň jeden z vás má zablácené čelo“.

Připomeňme, že ve hře zablácených dětí, se z p stala všeobecná znalost v okamžiku, kdy otec řekl p tak, aby všechny děti věděly,

- že platí p a
- že všechny děti to vědí.

Axiom C2.

$$\models C_G A \leftrightarrow E_G(A \wedge C_G A)$$

{ $C_G A$ je pevným bodem funkce $f(x) = E_G(A \wedge x)$ }

Následující odvozovací pravidlo dává způsob, jak ukázat, že v nějaké struktuře platí společná znalost.

Indukční pravidlo RC1

Pro všechny struktury M platí

je - li $M \models A \to E_G(B \wedge A)$, potom $M \models A \to C_G B$

Idea. Předpoklad indukčního pravidla dává možnost indukcí podle k dokázat, že formule

$$A \to E_G^{\,k}(B \wedge A) \quad \text{je validní pro každé } k.$$

Následující tvrzení dává sémantické odůvodnění charakteristiky operátoru E_G, Axiomu o pevném bodu a Indukčního pravidla.

Věta - shrnutí

Pro všechny struktury M, všechny neprázdné podmnožiny G množiny {1, 2, . . . , n} a libovolné formule A, B platí

$$(i) \quad M \models E_G A \;\Leftrightarrow\; \bigcap_{i \in G} K_i A$$

kde $\bigcap_{i \in G} K_i A$ označuje konjunkci všech $K_i A$ pro vš. i z G

$$(ii) \quad M \models C_G A \leftrightarrow E_G(A \wedge C_G A)$$

$$(iii) \; je-li \;\; M \models A \to E_G(B \wedge A), \;\; potom \;\; M \models A \to C_G B$$

Distribuované znalosti

charakterizují znalosti, které odpovídají tomu, když *„všichni agenti dají hlavy dohromady*“. Proto splňují stejné vlastnosti jako znalosti. Navíc upozorníme na dvě drobnosti.

$$\models D_{\{i\}}A \leftrightarrow K_i A$$

pro jednočlennou skupinu se distribuované znalosti shodují s tím, co agent ví.

Naopak, čím je skupina větší, tím větší jsou její distribuované znalosti.

$$Je-li \quad G \subseteq Q \ potom \ \models D_G A \rightarrow D_Q A$$

Anna a Bob

- **A**nna a **B**ob vědí, že organizátor vybere z osudí nějaké přirozené číslo ***n***, které napíše na čelo jednomu z nich a druhému napíše číslo „sousední“, tj. buď ***n*+1** nebo ***n*-1**. **Ani A**nna ani **B**ob neznají své číslo - vidí jen to partnerovo.
- Nakreslete odpovídající Kripkeho strukturu.
- Nechť **A** má na čele napsáno **3** a **B** zase **4**. Nakreslete odpovídající Kripkeho strukturu.
- Je společnou vlastností obou hráčů, že bylo vybráno číslo menší než 100?