Znalosti v multi-agentních systémech

Olga Štěpánková

Filozofové a logici se nejdřív soustředili na studium vlastností znalostí a odvozování v případě jediného individua.

Ale jádrem každé analýzy běžných situací

konverzace
obchodního vyjednávání
protokolu řízeného událostmi v distribuovaném prostředí

je interakce mezi (více) agenty

Naši agenti mohou být vyjednávači, komunikující roboti, vodiče, paměti nebo složité počítačové systémy

• Agent ve skupině musí brát v úvahu fakta, která jsou pravdivá v okolním světě,
• ale také znalosti ostatních agentů ve skupině.

Příklad. Kohout a pan Brouček.

Dean neví, jestli Nixon ví, že Dean ví, že Nixon ví, že McCord se vloupal do O´Brienovy kanceláře ve Watergate.

Většina lidí se rychle ztrácí v takto zahnízděné skupině znalostí, jedná-li se o „cizí“ prostředí.

Příklad.

Běžně se setkáváme se situací, kdy každý ve skupině ví určitý fakt.

*Každý řidič ví, že **červená je „stůj“ a zelená „volno“**. Tento fakt sám nestačí, abychom se na silnici cítili bezpečněji! **Proč?***

Potřebujeme si být jisti, že každý zná toto pravidlo a dodržuje ho. V některých aplikacích nestačí toto „dvoustupňové“ vědění.

Jsou situace, kdy je třeba uvažovat stav, *v němž současně* ***každý ví nějaký fakt F,***

každý ví, že ***každý ví F,***

každý ví, že každý ví, že ***každý ví*** *F atd.*

V tako vém případě říkáme, že skupina má **společnou znalost *F*.**

Společná znalost
• je nutná k porozumění v diskusi
• je podmínkou k dosažení dohody

Ušmudlané děti.

Výchozí stav *n* dětí si společně hraje, po určité době se *k* dětí ušmudlá na místě, které nemohou vidět, například na čele.
Přichází otec a říká: ***„Jeden z vás má špínu na čele“.***

{to je fakt známý každému dítěti, je-li *k > 1*}

Otec se opakovaně ptá ***„Ví někdo z vás, zda má špínu na čele ?“***

Mohou děti dospět k nějakému závěru?

Označíme tvrzení „*nejméně jeden z vás má špínu na čele*“ písmenem ***p***. Pokud $k > 1$, může se zdát, že otec tímto tvrzením neposkytl žádnou informaci.

Kdyby otec neřekl ***p***, ušmudlané děti nebudou nikdy schopny usoudit, že mají špínu na čele. Indukcí podle k se dá dokázat, že bez ohledu na situaci tj. na počet ušmudlaných dětí, všechny děti odpoví NE na prvních $q< k$ otázek.

k - 1 krát všechny děti odpoví NE, v ***k***-tém kole všechny ušmudlané děti odpoví ANO.

Ušmudlané děti.

Výchozí stav n dětí si společně hraje, po určité době se k dětí ušmudlá na místě, které nemohou vidět, například na čele.

Přichází otec a říká: ***„Jeden z vás má špínu na čele“.***

{to je fakt známý každému dítěti, je-li $k > 1$}

Otec se opakovaně ptá ***„Ví někdo z vás, zda má špínu na čele ?“***

Analýza: dříve než otec řekne ***p***, každý ví ***p*** ($k > 1$), ale není vždy pravda, že každý ví, že každý ví ***p*** .

První pokus o důkaz indukcí podle k.

Model znalostí

Kripkeho model možných světů.

Idea:
Kromě skutečného stavu světa existuje jistý počet možných stavů - „možné světy“.

S informacemi, které agent má, nemusí být schopen říci, který z možných světů popisuje skutečný stav.

Definice. Říkáme, že *agent zná fakt* ***p*** , jestliže ***p*** je pravdivé ve všech stavech, které agent pokládá za možné (vzhledem k informacím, které má).

Příklad.

Agent1 se prochází ulicemi Ústí nad Labem, kde je slunný den, ale nemá informaci o počasí v Humpolci.

Tedy ve všech světech, které Agent1 pokládá za možné, je v Ústí nad Labem slunný den.

Na druhé straně, protože agent neví jaké počasí je v Humpolci, v některých z jeho možných světů v Humpolci prší a v jiných je v Humpolci slunný den.

Agent1 tedy ví, že v Ústí nad Labem je slunný den, ale neví, je-li slunný den také v Humpolci.

Intuitivně, čím méně světů agent považuje za možné, tím je menší jeho nejistota a tím více toho agent ví.

Jestliže Agent1 získá ze spolehlivého zdroje informaci, že v Humpolci je slunný den, pak již nemusí dále uvažovat jako možné ty světy, ve kterých v Humpolci prší (je zataže-no, mlha a podobně).

Abychom mohli tyto myšlenky vyjádřit přesně, potřebujeme jazyk, který by dovolil vyjádřit pojmy týkající se znalostí jednoznačným způsobem.

Použijeme *jazyk výrokové modální logiky*.

Předpokládejme, že máme skupinu n agentů, které pojmenujeme $1, 2, \ldots, n$, kteří chtějí uvažovat o světě, který se dá popsat neprázdnou množinou prvotních výroků **Φ**, které budeme označovat
. $p, p', q, q', \ldots$

Prvotní výroky vyjadřují základní fakta o světě, například „v Humpolci prší“, „Mařenka je ušmudlaná“.

Abychom mohli vyjádřit tvrzení

„Karel ví , že prší v Humpolci"

rozšíříme jazyk o modální operátory

$$K_1, K_2, \dots, K_n$$

každý pro jednoho agenta.

Výraz $K_i p$ čteme *„ agent **i** ví **p**".*

K jazyku patří také základní výrokové spojky $\neg$ a $\wedge$ z nichž se dají ostatní spojky definovat.

Formule.

$p \in \Phi \Rightarrow p \in$ ***Formule***

$A, B \in$ ***Formule*** $\Rightarrow \neg A, (A \wedge B) \in$ ***Formule***

$A \in$ ***Formule*** $\& \ 1 \le i \le n \Rightarrow K_i A \in$ ***Formule***

Standardní zkratky z výrokové logiky

$A \vee B$ za $\neg(\neg A \wedge \neg B)$

$A \to B$ za $\neg A \vee B$

$A \leftrightarrow B$ za $((A \to B) \wedge (B \to A))$

true za $p \vee \neg p$ *false* za $\neg$*true*

{***p*** je pevně zvolená prvotní formule }

Příklad.

a)

$$K_1 K_2 p \wedge \neg K_2 K_1 K_2 p$$

Agent1 ví, že agent2 ví ***p*** , ale agent2 neví, že agent1 ví, že agent2 ví ***p*** .

b) možnost chápeme jako duální ke znalosti.
Agent1 považuje ***A*** za možné, jestliže agent 1 **neví** $\neg$ ***A*** , tj.

$$\neg K_1 \neg A$$

Uvažujme tvrzení

Dean neví zda Nixon ví, že Dean ví, že Nixon ví, že McCord se vloupal do kanceláře O´Briena ve Watergate.

Označíme-li Deana za agenta 1, Nixona za agenta 2 a p za výrok *„McCord se vloupal do kanceláře O´ Briena ve Watergate".*

Pak uvedené tvrzení lze zapsat takto

$$\neg K_1 \neg (K_2 K_1 K_2 p) \wedge \neg K_1 \neg (\neg K_2 K_1 K_2 p)$$

Sémantika (naší) modální logiky.

Kripkeho sémantika možných světů.

Kripkeho struktura M pro n agentů nad množinou prvotních formulí Φ je $(n+2)$-tice

$$(S, \pi, K_1, K_2, \dots, K_n)$$

kde S je množina možných světů nebo krátce stavů, π je interpretace stavů, která každému stavu s přiřazuje pravdivostní ohodnocení prvotních formulí z Φ, tedy

$$\pi(s) : \Phi \to \{true, \ false\}$$

a K_i jsou binární relace na S .

Na začátku našeho výkladu, budeme předpokládat, že tyto relace přípustnosti jsou **ekvivalence**. Potom

$$(s,t) \in K_i \Leftrightarrow (t,s) \in K_i$$

Znamená-li $(s,t) \in K_i$, že agent i ve stavu s považuje svět t za možný (přípustný), pak ze symetrie a tranzitivity plyne, že agent i má v s i t stejnou informaci o světě (stejnou množinu možných světů).

Stavy s a t jsou pro agenta *i* v tomto případě nerozlišitelné!, tento přístup se dá použít u řady aplikací.

Sémantika možných světů.

Budeme definovat pojem **(M, s)** |= ***A***, který čteme „formule *A platí ve struktuře M a stavu s*“ nebo „ *A je splněna v (M, s)*“. Postupujeme indukcí podle struktury A.

$(i)\quad (M,s) \models p \quad$ *právě když* $\quad \boldsymbol{\pi}(s)(p) = true \ \{p \in \boldsymbol{\Phi}\}$

$(ii)\quad (M,s) \models \neg A \quad$ *právě když* $\quad (M,s) \not\models A$

$(iii)\quad (M,s) \models A \wedge B$ *právě když* $\quad (M,s) \models A$ *a* $(M,s) \models B$

$(iv)\quad (M,s) \models K_i A \quad$ *právě když* $\quad (M,t) \models A$

pro všechna $t, (s,t) \in K_i$

Kripkeho struktury lze zobrazit jako ohodnocené orientované grafy.

Uzly grafu jsou stavy $s \in S$. Uzly jsou ohodnoceny množinou prvotních formulí, které v s platí.

Orientované hrany ohodnocujeme množinami agentů, ohodnocení hrany z uzlu ***s*** do ***t*** obsahuje index ***i***, jestliže $(s,t) \in K_i$.

Příklad.

Nechť $\Phi = \{p\}$ a $n = 2$, tedy náš jazyk má jednu prvotní formuli p a existují dva agenti.

Uvažujme Kripkeho strukturu

$$M = (S, \boldsymbol{\pi}, K_1, K_2)$$

kde

(i) $S = \{s, t, u\}$

(ii) ***p*** je pravdivé ve stavech s a u, ne však v t. Tedy

$\pi(s)(\boldsymbol{p}) = \pi(u)(\boldsymbol{p}) = true$ a $\pi(t)(\boldsymbol{p}) = false$

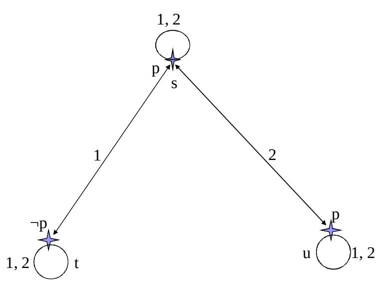
(iii) agent1 neumí rozlišit stav s od t, takže

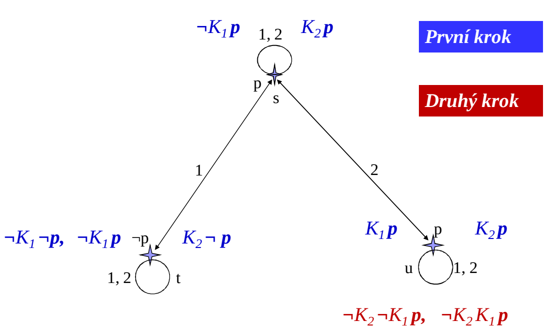
$$K_1 = \{(s,s),(s,t),(t,s),(t,t),(u,u)\}$$

agent2 neumí rozlišit stav s od u, tedy

$$K_2 = \{(s,s),(s,u),(u,s),(t,t),(u,u)\}$$

Situaci znázorníme následujícím grafem, který popisuje relace K_i.





Je-li p výrok „ v Ústí nad Labem je slunný den“, potom ve stavu s je v Ústí nad Labem slunný den, ale algent1 o tom neví, protože ve stavu s považuje za možné oba stavy s a t.

Agent1 je si vědom, že s a t jsou dva různé stavy, to co chceme říci je, že agent1 nemá dostatečné informace, aby rozlišil stavy s a t.

Agent2 ve stavu s, ví že v Ústí nad Labem je slunečno, protože ve stavu s považuje za možné jen stavy s a u a v obou p platí.

Agent2 ví i ve stavu t jaký je skutečný stav, že není slunečno.

Z toho plyne, že ve stavu s, agent1 ví, že agent2 ví v zda v Ústí nad Labem je slunný den nebo ne v obou stavech, které agent1 považuje za možné ve stavu s, jmenovitě je stavech s a t, agent2 zná skutečný stav věcí v obou z nich.

Tedy ačkoliv agent1 neví ve stavu s skutečnou situaci, ví že agent2 zná skutečnou situaci ve stavu s.

V kontrastu k tomu, i když agent2 ve stavu s ví, že v Ústí nad Labem je slunný den, neví , že agent1 nezná tento fakt {v jednom světě, který agent2 považuje za možný, jmenovitě u agent1 ví že v Ústí je slunečno, ale ve druhém možném světě agenta2, jmenovitě s agent1 tento fakt neví}.

Tato komplikovaná úvaha může být shrnuta do jediného sémantického tvrzení

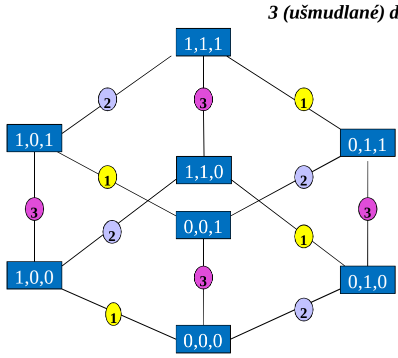
$$(M,s) \models p \wedge \neg K_1 p \wedge K_2 p \wedge K_1(K_2 p \vee K_2 \neg p) \wedge \neg K_2 \neg K_1 p$$

Protože ve stavech s a u má naše jediná prvotní formule stejné ohodnocení, zdálo by se, že je možné jeden z nich vynechat. To ale není možné.

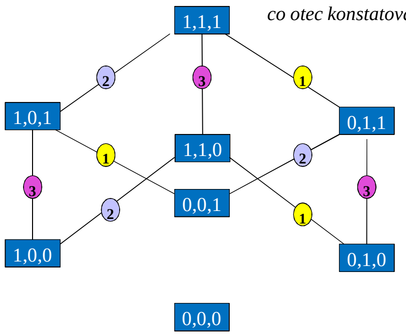
Stav není určen jen pravdivostním ohodnocením, ale také relacemi mezi možnými světy.

Kdyby agent1 považoval ve stavu s za možný stav u místo stavu t, věděl by jaká je situace ve stavu s.

Kripkeho struktura pro 3 (ušmudlané) děti



Kripkeho struktura pro 3 ušmudlané děti *poté, co otec konstatoval* ***p***



Všeobecné a distribuované znalosti

K vyjádření těchto pojmů přidáme do jazyka tři modální operátory

E_G {"*každý ve skupině G ví*"}

C_G {" *je to všeobecná znalost mezi agenty v G*"}

D_G {" *je to distribuovaná znalost mezi agenty v G*"}

pro každou neprázdnou podmnožinu G množiny [1, 2, . . . , n]. Je-li A formule, potom $E_G A, C_G A$ a $D_G A$ jsou také formule.

Příklad.

$K_3 \neg C_{[1,2]} p$ agent3 ví, že p není všeobecná znalost mezi agenty 1 a 2.

$D_G q \wedge \neg C_G q$ q je distribuovaná znalost, ale není to znalost všeobecná.

Není těžké definovat sémantiku těchto operátorů. Nejprve definujme iteraci operátoru E_G.

$E_G^0 A \equiv A$ $\qquad E_G^{n+1} A \equiv E_G E_G^n A$

Definujeme

$$(M,s) \models E_G A \iff (M,s) \models K_i A \text{ pro všechna } i \in G$$

$$(M,s) \models C_G A \iff (M,s) \models E_G^k A \text{ pro všechna } k$$

Oba pojmy mají zajímavou grafovou interpretaci, je-li G neprázdná podmnožina agentů, říkáme, že stav ***t*** je G-dosažitelný ze stavu ***s*** v ***k*** krocích, jestliže existuje posloupnost stavů

$$s \equiv s_0, s_1, \ldots, s_k \equiv t$$

taková, že pro každé $j, 0 \leq j < k$ existuje $i \in G$ takové, že $(s_j, s_{j+1}) \in K_i$. Říkáme, že *t je G-dosažitelné z s*, je-li G-dosažitelné po konečném počtu kroků.

Lemma.

(i) $(M,s) \models E_G^k A \iff (M,t) \models A$ *pro každé t, G − dosažitelné v k krocích*

(ii) $(M,s) \models C_G A \iff (M,t) \models A$ *pro každé t, G − dosažitelné z s.*

Důkaz.

(i) se dokáže indukcí podle k, (ii) plyne z (i).

Obě tvrzení se dokáží pro libovolné relace K, žádná jejich vlastnost se nepředpokládá.

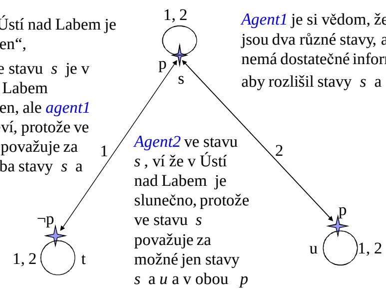
Distribuované znalosti.

Vraťme se k našemu modelu z Ústí nad Labem.

p = „v Ústí nad Labem je slunný den“,

potom ve stavu s je v Ústí nad Labem slunný den, ale *agent1* o tom neví, protože ve stavu s považuje za možné oba stavy s a t.

Agent1 je si vědom, že s a t jsou dva různé stavy, ale nemá dostatečné informace, aby rozlišil stavy s a t.



Agent2 ve stavu s, ví že v Ústí nad Labem je slunečno, protože ve stavu s považuje za možné jen stavy s a u a v obou p platí.

Ve stavu s *agent1* považuje za možné oba stavy s i t, ale ne stav u. Přitom *agent2* považuje za možné stavy s a u, zatímco stav t ne.

Kdo by uměl využít znalosti obou agentů, ten by věděl, že možný je jenom stav s: *agent1* má dost znalostí, aby mohl vyloučit stav u a *agent2* by ze stejného důvodu vyloučil stav t.

Obecně, kombinujeme znalosti agentů ze skupiny G, abychom vyloučili všechny světy, které některý agent považuje za nemožné.

Tomu odpovídá průnik relací K. Definujeme *distribuovanou znalost* D_G vztahem

$$(M,s) \models D_G A \iff (M,t) \models A \text{ pro každé } t, \; (s,t) \in \bigcap_{i \in G} K_i$$

Je-li počet ušmudlaných $k = 2$, není těžké ukázat,že neplatí, že každý ví, že každý ví p .

Naopak, je-li $k = 3$, tvrzení že každý ví, že každý ví p , ale neplatí, že každý ví,že každý ví, že každý ví p. (3 krát)

Označme

$E^k p$ *tvrzení* (*každý ví, že*)^k *platí* p

$C\,p$ *tvrzení, že* p *společná znalost*

Cvičení Je-li právě k dětí ušmudalných, dříve než otec promluví, platí $E^{k-1} p$ ale neplatí $E^k p$.

Otcovo prohlášení mění stav znalostí dětí z $E^{k-1} p$ na $C\,p$.

Zatím jsme pracovali s *výrokovou modální logikou*.

Nemáme zde kvantifikaci prvního řádu univerzální ani existenční, proto nemůžeme popsat výroky: „Mařenka umí vyjmenovat (zná jménem) všechny krajské hejtmany“.

V *predikátové modální logice* bychom napsali

$(\forall x)(Kraj(x) \to (\exists y)(K_{Mařenka} Krajský\_hejtman(x, y))$

V dalším zůstaneme u výrokové modální logiky, která postačí pro naše účely a vyhneme se tak komplikovaným situacím, které přináší predikátová modální logika.