Syntaxe a sémantika BDI logiky

František Zbořil
zborilf@dcse.fee.vutbr.cz

Abstrakt: V současnosti je v oblasti umělé inteligence věnována velká pozornost výzkumu agentních systémů. Pro účel studie těchto systémů je snaha o vytvoření formálního modelu. Jednou z možností vytvoření takových formálních modelů je model založený na BDI logice . Text pojednává o formalismech, ze kterých BDI logika vychází, popisuje její syntaxi a ukazuje některé z axiomů, na kterých je tato logika postavena.

1. Úvod

Během posledních dvaceti let probíhá intenzivní výzkum v oblasti distribuovaných systémů. Podnětem se stal bezesporu i razantní nástup celosvětových počítačových sítí, jenž dal podnět k řešení problematiky systémů pro distribuované výpočty a distribuovaná řízení, jejich komunikaci, příp. bezpečnost provozu. Společně s těmito tématy se distributivní přístup přesunul i do oblastí jako je umělá inteligence, kde se stal jedním z nejvíce zkoumaným tématem tzv. systémy založené na agentech. Umělá inteligence rozdělená na víceméně samostatné podoblasti, jako jsou například řešení úloh, plánování, počítačové vidění či reprezentace znalostí byla často terčem kritiky, která poukazovala na to, že myšlenka nejprve samostatně vyřešit jednotlivé podoblasti a následně na jejich základě sestavit funkční, komplexní systém je mylná. Kritici tohoto přístupu [Brooks87] naopak argumentují tím, že je nutné postupovat od jednoduchých systémů k složitějším a mít přitom v každé fázi vývoje funkční systém schopný pracovat v reálném prostředí a provádět reálné operace. Cílem tohoto článku je přiblížit dnes pravděpodobně nejpopulárnější přístup k formalizaci systémů založených na agentech – BDI architektuře. Kapitola 2 uvádí formalismy, z kterých BDI logika vychází, kapitola 3 je zaměřena přímo na BDI logiku. Jsou zde uvedeny základní axiomy a inferenční pravidla logiky. Dále jsou zmíněna pravidla pro konzistentnost této logiky a vztahy mezi jednotlivými složkami – mentálními postoji agenta.

2. Modelování systémů

Při studiu systému vytváříme jeho formální modely, na kterých máme možnost zkoumat jeho chování, jeho stavy a jeho reakce na podněty. V takovém případě tedy máme stav systému, které můžeme popsat výroky výrokové logiky. Těmto výrokům jsou tedy v každém ze stavů přiřazeny pravdivostní hodnoty a chování systému, resp. jeho reakce na události je popsána grafem. Během procesu simulace se stav systému mění a s ním se mění i platnost výroků. Následující odstavce popíší formální struktury, ze kterých BDI logika vychází.

Kripkeho struktura

Mějme množinu atomických výroků AP. Potom Kirpkeho strukturu definujeme jako čtveřici $M=(S,S_0,R,L)$, kde

- S je konečná množina stavů
- $S_0 \subseteq S$ je množina počátečních stavů

- $R \subseteq SxS$ je totální tranzitivní relace. Totální relace je taková, že pro každý stav $s \subseteq S$ existuje stav $s' \subseteq S$ a platí $R(s,s')$
- $L:S \rightarrow 2^{AP}$

Tedy v každém stavu je přiřazena atomickým výrokům pravdivostní hodnota. Cestou (path) systému je posloupnost stavů z *s*, tedy

$\pi = s_0\, s_1\, s_2 \ldots$, pro $i \geq 0$ platí $R(s_i, s_{i+1})$

Společně s relacemi mezi stavy Kripkeho struktura umožňuje modelovat chování nedeterministického konečného systému. Tato je využívána například při modelování paralelních systémů nebo pro verifikaci. Jejím rozšířením jsou CTL* a CTL logiky.

CTL* a CTL logika

CTL* (The Computation Tree Logic) tvoří Kripkeho strukturu s jedním počátečním stavem. Dále zavádí boolovské operátory k atomickým výroků. Formule jsou tvořeny temporalními operátory (temporal operators) a kvantifikátory cesty (path quantificators). Kvantifikátory cesty jsou ***A*** *(*pro všechny výpočetní cesty) a ***E*** (pro některou z výpočetních cest). Temporálními operátory jsou: ***X*** (v příštím kroce), ***F*** (v budoucnosti), ***G*** (vždy, globálně), ***U*** (dokud)

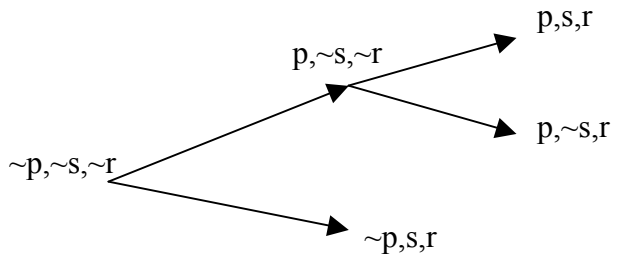
Stavové formule a formule cesty jsou pro CTL logiku definovány následovně:

- (S1) Jestliže $p \in AP$, potom *p* je stavová formule
- (S2) Jestliže *f* a *g* jsou stavové formule, potom $\neg f$, $f \vee g$ a $f \wedge g$ jsou stavové formule
- (S3) Jestliže *f* je formule cesty, potom ***Ef*** a ***Af*** jsou stavové formule

Pro formule cesty platí následující pravidla:

- (P1) Jestliže *f* je stavová formule, potom *f* je i formulí cesty
- (P2) Jestliže *f* a *g* jsou formule cesty, potom i $\neg f$, $f \vee g$, $f \wedge g$, jsou formule cesty
- (P3) Jestliže *f* a *g* jsou formule cesty, potom i ***Xf***, ***Ff***, ***Gf*** a ***fUg*** jsou formule cesty

Na následujícím obrázku budou ukázány některé stavové formule a formule cesty



obrázek 1 Graf CTL stromové struktury

Uvažujeme-li stav, který je na tomto obrázku kořenem, potom stavovými formulemi v tomto místě jsou například ~p ^ ~r, **EX**p, **AF**r, a formulemi cesty jsou **X**b, **F**r, **X**b^**F**r.

Sémantiku CTL logiky můžeme definovat pomocí následujících axiomů. Máme dvě stavové formule f1, f2 a dvě formule cesty g1, g2. Zápis M,s |= f značí, že M je Kripkeho struktura a výrok *f* je splněn ve stavu *s*. Obdobně pro cesty zápis M,π |= f znamená, že formule f je v Kripkeho struktuře M splněna ve všech stavech cesty π. Značení π^i je použito pro příponu (suffix) cesty π začínající ve stavu s_i. Potom pro jednotlivé operátory platí:

(S1) $M,s \models p \quad \Leftrightarrow \quad p \in L(s)$

(S2) $M,s \models \neg f_1 \quad \Leftrightarrow \quad M,s \not\models f_1$
$M,s \models f_1 \vee f_2 \quad \Leftrightarrow \quad M,s \models f_1$ nebo $M,s \models f_2$
$M,s \models f_1 \wedge f_2 \quad \Leftrightarrow \quad M,s \models f_1$ a zároveň $M,s \models f_2$

(S3) $M,s \models \mathbf{E}\, g_1 \quad \Leftrightarrow \quad$ existuje cesta π ze stavu *s* taková, že $M, \pi \models g_1$
$M,s \models \mathbf{A}\, g_1 \quad \Leftrightarrow \quad$ pro každou cestu π ze stavu *s* platí, že $M, \pi \models g_1$

(P1) $M,\pi \models f_1 \quad \Leftrightarrow \quad$ s je prvním stavem cesty π a $M,s \models f_1$

(P2) $M,\pi \models \neg g_1 \quad \Leftrightarrow \quad M,\pi \not\models g_1$
$M,\pi \models g_1 \vee g_2 \quad \Leftrightarrow \quad M, \pi \models g_1$ nebo $M, \pi \models g_2$
$M,\pi \models g_1 \wedge g_2 \quad \Leftrightarrow \quad M, \pi \models g_1$ a zároveň $M, \pi \models g_2$
(P3) $M,\pi \models \mathbf{X}\, g_1 \quad \Leftrightarrow \quad M, \pi^1 \models g_1$

$M,\pi \models \mathbf{F}\, g_1 \quad \Leftrightarrow \quad$ existuje konstanta $k \geq 0$ a $M, \pi^k \models g_1$
$M,\pi \models \mathbf{G}\, g_1 \quad \Leftrightarrow \quad$ pro všechna $k \geq 0$ platí $M, \pi^k \models g_1$
$M,\pi \models g_1\, \mathbf{U}\, g_2 \quad \Leftrightarrow \quad$ existuje konstanta $k \geq 0$ a $M, \pi^k \models g_2$ a pro všechna $0 \leq j \leq k$ platí $M, \pi^j \models g_1$
$M,\pi \models g_1\, \mathbf{R}\, g_2 \quad \Leftrightarrow \quad$ pro všechna $k \geq 0$ platí, že pokud pro všechna $i < k$ $M, \pi^i \not\models g_1$, potom $M, \pi^k \models g_2$

Zřejmě se nejedná o nejmenší možný soubor logických operátorů. Stejně jako pomoci De Morganových zákonů můžeme vyjádřit konjunkci pomocí negace a disjunkce, lze i některé z temporálních operátorů nahradit vhodným zápisem ostatních. Jako příklad můžeme uvést následující vztahy:

$$f\, \mathbf{R}\, g \equiv \neg(\neg f\, \mathbf{U}\, \neg g)$$
$$\mathbf{F}\, f \equiv \text{True}\, \mathbf{U}\, f$$
$$\mathbf{G}\, f \equiv \neg \mathbf{F} \neg f$$
$$\mathbf{A}(f) \equiv \neg \mathbf{E}(\neg f)$$

Kompletní sada axiomů a inferenčních pravidel pro CTL bude uvedena v další kapitole při axiomizaci BDL logiky.

CTL logika je podmnožina CTL* logiky, která omezuje formule cesty a uvažuje pouze takové, které vyhovují definici, že

- (P0) pokud *f* a *g* jsou stavové formule, potom **X** f, **F** f, **G** f, f **U** g a f **R** g jsou formulemi cesty.

Tedy například formule f,g jsou stavovými formulemi a **X**f $\vee$ **F**g patří do CTL* logiky, ale do logiky CTL nepatří. V literatuře pojednávající o BDI logice a BDI architekturách se vychází právě z formu CTL logiky.

3. BDI Logika

Následující kapitola uvádí rozšíření CTL logiky tak, aby jím bylo možné modelovat systém založený na agentech. Chování agentů vychází z jejich mentálních stavů, jimiž jsou jejich představa o prostředí, ve kterém se nacházejí (Beliefs), jejich tužby (Desires) a záměry (Intentions). Tyto pojmy mají blízko ke klasickým pojmům z oblastio umělé inteligence, jako jsou znalost (knowledge) nebo cíl (goal), ale tyto nepopisují zcela to samé. Agent má nějakou znalost, ale touto znalostí si nemusí být zcela jist, proto místo znalosti se používá pojem představa, domněnka. Místo pojmu cíl používáme pojmu tužba. Tužby se mohou měnit s tím, jak se mění agentovy představy o prostředí. Pojmem záměr rozumíme krátkodobý cíl, zpravidla podúlohu vykonávanou za účelem realizace některé z tužeb. Je zřejmě vidět, jak vznikl název pro BDI logiku. Jde o zkratku vzniklou ze zde popisovaných vlastností. Aby tyto vlastnosti mohli být zahrnuty v modelu systému s agenty, je o jejích specifikaci rozšířena popsaná CTL logika. Potom tedy rozšíříme CTL o tři modální operátory **BEL**(f), **DES**(f) a **INT** (f) pro domněnky, tužby a záměry. Tedy formule definujeme jako (S1) – (S3), (P0) a navíc

- (S4) pokud *f* je stavová formule, potom i **BEL**(*f*), **DES**(*f*) a **INT**(*f*) jsou stavovými formulemi

Definujme Kripkeho strukturu M=<W,{S_w: w$\in$W},{R_w:w$\in$W},L,*B*,*D*,*I*>, kde W je množina všech možných světů, S_w je množina stavů pro každý ze světů, R_w je totální binární relace, tedy $R_w \subseteq S_w$ x S_w , L přiřazuje pravdivostní hodnotu atomickým formulím každému ze světů v každém ze stavů s$\in S_w$,.a *B*, *D*, *I* jsou relace nad světy, tedy B$\subseteq$W x W.

Pojmem svět rozumíme stromovou strukturu, s lineární minulostí a rozvětvenou budoucností. Uzly tvoří časové okamžiky, ve kterých je několik možných událostí. Z definice plyne, že v každém uzlu – časovém okamžiku jsou přiřazeny pravdivostní hodnoty atomickým formulím. Vyhodnocení formule tedy závisí na světu a stavu, ve kterém se nachází. Relace *B* značí vztahy mezi světy ve smyslu, že agent věří, které ostatní světy jsou možné. Obdobně lze chápat relace pro tužby a intence *D*, *I*.

Sémantika pro BDI logiku je obdobná jako pro logiku CTL s tím rozdílem, že pro stavy a cesty chápeme v rámci nějakého ze světů, tedy například definicí sémantiky pro (S1) by bylo

(S1) $M, w_s \models p \quad \Leftrightarrow \quad p \in L(w,s)$

Pro nové operátory zavedeme sémantiku následující

(S4) $M, w_s \models$ **BEL**(*f*) $\Leftrightarrow \forall v, (w,s,v) \in B$; $M, v_s \models f$
$M, w_s \models$ **DES**(*f*) $\Leftrightarrow \forall v, (w,s,v) \in D$; $M, v_s \models f$
$M, w_s \models$ **INT**(*f*) $\Leftrightarrow \forall v, (w,s,v) \in I$; $M, v_s \models f$

Tedy formule *f* je agentem ‘věřena’, pokud je platná ve všech světech (tedy světech, které jsou v relaci *B*) dostupných agentovi ze stavu s světa w. Obdobně je tomu i v případech tužeb a intencí. Vztahy mezi doměnkami, tužbami a intencemi budou uvedeny v závěrečném odstavci. Nyní bude uvedena soustava axiomů a inferenčních pravidel. Nejprve jsou uvedeny axiomy pro CTL logiku

(CTL1)	Zahrnuje všechny axiomy výrokové logiky		
(CTL2)	**EF**f	≡	**E**(*true* **U** *f*)
(CTL2b)	**AG**f	≡	¬**EF**¬f
(CTL3)	**AF**f	≡	**A**(*true* **U** *f*)
(CTL3b)	**EG***f*	≡	¬**AF** ¬*f*
(CTL4)	**EX**(*f* ∨ g)	≡	**EX**f ∨ **EX**g
(CTL5)	**AX**f	≡	¬**EX**¬f
(CTL6)	**E**(*f* **U** *g*)	≡	*g* ∨ (*f* ∧ **EXE**(*f* **U** *g*))
(CTL7)	**A**(*f* **U** *g*)	≡	*g* ∨ (*f* ∧ **AXA**(*f* **U** *g*))
(CTL8)	**EX**true ∧ **AX**true		
(CTL9)	**AG**(*h* → (¬*f* ∧ **EX***h*)) → (*h* → ¬**A** (*g* **U** *f*))		
(CTL9b)	**AG**(*h* → (¬*f* ∧ **EX***h*)) → (*h* → ¬**AF***f*))		
(CTL10)	**AG**(*h* → (¬*f* ∧ (g ⊃ **AX***h*))) → (*h* → ¬**E** (*g* **U** *f*))		
(CTL10b)	**AG**(*h* → (¬*f* ∧ **AX***h*)) → (*h* → ¬**EF***f*)		
(CTL11)	**AG**(*f* → *g*) → (**EX***f* → **EX***g*)		
(CTL-Gen)	pokud |- *f*, potom |- **AF**f		
(MP)	pokud |- *f* a |- *f* → *g* , potom |- *g*		

Axiomy pro domněnky, tužby a záměry zahájíme K axiomy modální logiky. Tyto axiomy vyjadřují, že pokud agent věří *f,* a platí, že *f* → *g,* potom agent věří také *g.* Tento axiom uvedeme i ro případy tužeb a intencí.

(B-K) **BEL**(*f*) ∧ **BEL**(*f* → *g*) → **BEL**(*g*)
(D-K) **DES**(*f*) ∧ **DES**(*f* → *g*) → **DES**(*g*)
(I-K) **INT**(*f*) ∧ **INT**(*f* → *g*) → **INT**(*g*)

Následující pravidla (nazývána také generalizační pravidla) deginují pro jednotlivé složky mentálního postoje agenta skutečnost, že pokud je nějaká formule platná, potom je taková formule i věřena, chtěna i součástí záměru. Tedy platí následující axiomy

(B-Gen)	pokud |- *f* , potom **BEL**(*f*)
(D-Gen)	pokud |- *f* , potom **DES**(*f*)
(I-Gen)	pokud |- *f* , potom **INT**(*f*)

D axiom je axiomem modální logiky. Jeho význam je takový, že pokud se agent domnívá o nějaké skutečnosti že je pravdivá, potom není jeho domněnkou to, že by formule byla nepravdivou. Tento axiom přijmeme i pro tužby a domněnky. Axiomy 4 a 5 vyjadřují, že pokud agent věří v pravdivost formule *f*, potom také této skutečnosti věří tomu. Stejně tak platí i pro případ, že agent dané formuli nevěří.

(B-D) $\mathbf{BEL}(f) \rightarrow \neg\mathbf{BEL}(\neg f)$
(B-4) $\mathbf{BEL}(f) \rightarrow \mathbf{BEL}(\mathbf{BEL}(f))$
(B-5) $\neg\mathbf{BEL}(f) \rightarrow \mathbf{BEL}(\neg\mathbf{BEL}(f))$
(D-D) $\mathbf{DES}(f) \rightarrow \neg\mathbf{DES}(\neg f)$
(I-D) $\mathbf{INT}(f) \rightarrow \neg\mathbf{INT}(\neg f)$

Těmito axiomy a pravidly je tedy tvořen systém BDI logiky. Analýza takového systému je provedena např. v [1], kde jsou studovány rozšíření tohoto systému, konzistentnost systému, splnitelnost formulí apod.

Pravidla pro konzistenci BDI systému

Dále uvedeme některá pravidla vymezující konzistentnost formulí. Pokud je atomická formule pravdivá v nějakém stavu *s* světa *w*, potom nemůže být zároveň v tomot světě nepravdivá. Formálně zapsáno

(PC0) pokud $f \in \mathrm{L(w,s)}$, potom $\neg f \notin \mathrm{L(w,s)}$

Jiným příkladem pravidel pro konzistenci jsou tzv. pravidla pro lokální konzistenci. Ty stanoví, že pokud v některé nebo ve všech následujících stavech bude zaručena platnost nějaké stavové formule, potom formule v některém, resp. všech následujících stavech je pravdivá.

(LC0) Pokud $\mathbf{AX}f \in \mathrm{L(w,s)}$, potom pro $\forall$ následníky uzlu *s* uzly *t* platí $f \in \mathrm{L(w,t)}$
(LC1) Pokud $\mathbf{EX}f \in \mathrm{L(w,s)}$, potom $\exists$ následník uzlu *s* uzel *t* a platí $f \in \mathrm{L(w,t)}$

Poslední skupinou pravidel pro konzistenci, kterou si uvedeme jsou BDI – konzistentní pravidla. Vychází z již definované skutečnosti, že pokud je formuli *f* věřeno, tedy platí $\mathbf{BEL}(f)$, potom musí být formule *f* pravdivá ve všech světech, jež jsou v daném stavu daného světa v relaci *B*. Tedy musí platit, že

(BC0) pokud $\mathbf{BEL}(f) \in \mathrm{L(w,s)}$ a $(w,s,v) \in B$, potom $f \in \mathrm{L(v,s)}$
(BC1) pokud $\neg\mathbf{BEL}(f) \in \mathrm{L(w,s)}$ potom $\exists v$, $(w,s,v) \in B$, a $\neg f \in \mathrm{L(v,s)}$

Obdobná BDI-konzistentní pravidla jsou definována i pro zbývající dvě složky mentálního postoje agenta – tužby a intentce.

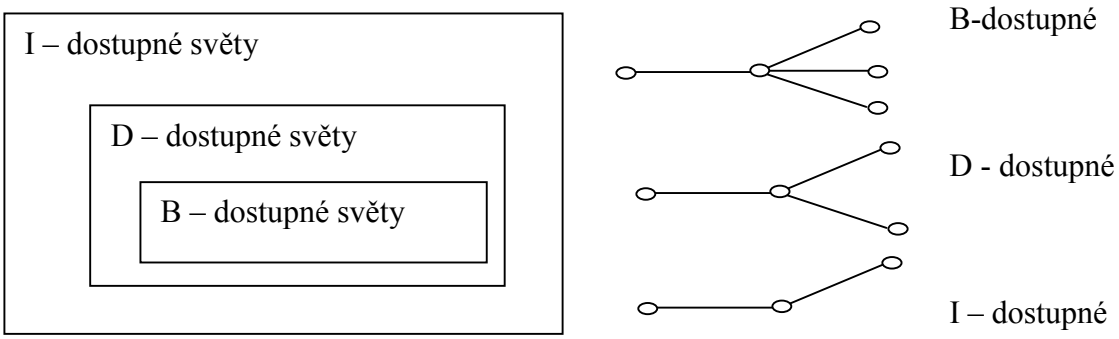
Vztah mezi jednotlivými složkami mentálního postoje agenta

V následujících odstavcích bude diskutován vztah mezi B-dosažitelnými, D- dosažitelnými a I-dosažitelnými světy a mezi strukturami jednotlivých světů. Dosažitelné světy z nějakého uzlu nějakého světa tvoří množinu. Dvě množiny S1 a S2 mohou být ve vztahu takovém, že S1 je podmnožinou S2, S2 je podmnožinou S1, S1 a S2 jsou stejné, mají nějaký průnik, nebo je jejich průnikem množina prázdná. Jednotlivé světy, jak již bylo uvedeno, mají stromovou strukturu, tedy dva světy *w* a *v* mohou být ve vztahu takovém, že *w* je podsvětem světa *v*, nebo naopak *v* je podsvětem světa *w*, oba světy mají stejnou strukturu, nebo se jejich struktura

liší. Svět *w* je podsvětem světa *v*, pokud struktura světa *v* je podstromem struktury světa *w* a ve všech uzlech mají oba světy stejné pravdivostní ohodnocení formulí.

Je patrné že existuje řada kombinací vztahů mezi jednotlivými složkami mentálního postoje, přičemž pouze některé z nich jsou brány jako vhodné pro modelování chování agenta. Nejčastěji bývají zmiňovány vztahy označované jako silný realismus, realismus a slabý realismus. Jako příklad uvedeme první z nich – silný realismus, kde množina B – dostupných světů je podmnožinou množiny D – dostupných světů a ta je podmnožinou množiny I – dostupných světů.

Silný realismus



obrázek 2 Vztah mezi BDI dostupnými světy v případě silného realismu

Silný realismus omezuje vztah mezi jednotlivými složkami tak, že pokud agent má nějaký záměr, potom existuje alespoň jeden D-dostupný svět, který touží vykonat tento záměr a dále ve vztahu k B-dostupným světům existuje alespoň jedna cesta v B-dostupném světě, kterou agent věří, že je jeho tužba uskutečnitelná.

4. Závěr

Byl ukázán jeden z přístupů k formální specifikaci agentních systémů Text si kladl za cíl ukázat základní myšlenky přístupu k vytváření formálního modelu těchto systémů pomocí rozšířené CTL logiky o mentální postoje agenta, které zahrnují jeho domněnky o okolním světě, tužby, vyjadřující cíle, jenž chce dosáhnout a záměry agenta. BDI logika je v současnosti nejznámější a nejpopulárnějším přístupem v této oblasti.

5. Literatura

[1] Rao, A.,S.: *Modeling Rational Agents within a BDI-Architecture*, Proceedings of the Second International Conference on principles of Knowledge Representation and Reasoning, San Mateo, 1991
[2] Rao, A.,S.: *Formal Models and Decision Procedures for Multi-Agent Systems*
[3] Mařík, V., Štěpánková O., Lažanský, J.: *Umělá inteligence (3)*, Academia, Praha 2000, ISBN 80-200-0472-6
[4] Clarke, E.,M. jr. Grumberg, O., Peled D.,A.: *Model Checking*, MIT Press 1999, ISBN 0-262-03270-8