Математическая логика и логическое программирование

mk.cs.msu.ru → Лекционные курсы
→ Математическая логика и логическое программирование (3-й поток)

Блок 49

Эпистемические логики

Лектор:
Подымов Владислав Васильевич
E-mail:
valdus@yandex.ru

ВМК МГУ, 2024/2025, осенний семестр

Эпистемическая логика (логика знаний) — это разновидность модальной логики, в которой модальность $\Box$ означает «я знаю», а $\Diamond$ — «я допускаю»

Так как эпистемическая логика является модальной, то в ней справедливы все законы модальной логики

Но смыслом модальностей могут определяться и другие законы

Если «я» — **идеальный познающий субъект**, то совокупность моих знаний и допущений должна подчиняться, помимо общих законов модальных логик, *как минимум* таким законам:

- мои знания верны
  $\Box\varphi \to \varphi$ (закон адекватности знания)
- мне известно, что именно я знаю
  $\Box\varphi \to \Box\Box\varphi$ (закон позитивной интроспекции)
- мне известно, что именно я **не** знаю
  $\neg\Box\varphi \to \Box\neg\Box\varphi$ (закон негативной интроспекции)

Двуместное отношение $\mapsto$ на множестве W

- рефлексивно, если для любого элемента w множества W верно $w \mapsto w$

- симметрично, если для любых элементов w_1, w_2 множества W справедлива импликация
$$w_1 \mapsto w_2 \quad \Rightarrow \quad w_2 \mapsto w_1$$

- транзитивно, если для любых элементов w_1, w_2, w_3 множества W справедлива импликация
$$w_1 \mapsto w_2 \mapsto w_3 \quad \Rightarrow \quad w_1 \mapsto w_3$$

Утверждение. **Для любой шкалы Крипке $\mathcal{F} = (W, \mapsto)$ верно:**

$\mathcal{F} \models \Box\varphi \to \varphi$ **верно для любой формулы** φ

$\Leftrightarrow$

отношение $\mapsto$ рефлексивно

Доказательство.

($\Rightarrow$) Пусть отношение $\mapsto$ нерефлексивно

Тогда существует мир w, такой что $w \not\mapsto w$

Рассмотрим переменную p и модель Крипке $\mathcal{I} = (W, \mapsto, L)$, такие что:

- $p \notin L(w)$
- Для любой w-альтернативы w' верно $p \in L(w')$

По выбору w, модель $\mathcal{I}$ задана корректно

При этом $\mathcal{I}, w \models \Box p$ и $\mathcal{I}, w \not\models p$, а значит, и $\mathcal{I}, w \not\models \Box p \to p$

Следовательно, $\mathcal{F} \not\models \Box p \to p$

Утверждение. **Для любой шкалы Крипке $\mathcal{F} = (W, \mapsto)$ верно:**

$\mathcal{F} \models \Box\varphi \to \varphi$ **верно для любой формулы** φ

$$\Leftrightarrow$$

отношение $\mapsto$ рефлексивно

Доказательство.

($\Leftarrow$) Пусть отношение $\mapsto$ рефлексивно

Предположим от противного, что существуют формула φ, модель Крипке $\mathcal{I} = (W, \mapsto, L)$ и её мир w, такие что $\mathcal{I}, w \not\models \Box\varphi \to \varphi$

По семантике $\to$ и $\Box$:

1. $\mathcal{I}, w \not\models \varphi$
2. Для любой w-альтернативы w' верно $\mathcal{I}, w' \models \varphi$

По рефлексивности $\mapsto$, мир w является w-альтернативой

Значит, $\mathcal{I}, w \models \varphi$, что *противоречит* первому пункту ▼

Утверждение. **Для любой шкалы Крипке $\mathcal{F} = (W, \mapsto)$ верно:**

$\mathcal{F} \models \Box\varphi \to \Box\Box\varphi$ **верно для любой формулы** φ

$$\Leftrightarrow$$

отношение $\mapsto$ транзитивно

Утверждение. **Для любой шкалы Крипке $\mathcal{F} = (W, \mapsto)$ с рефлексивным и транзитивным отношением $\mapsto$ верно:**

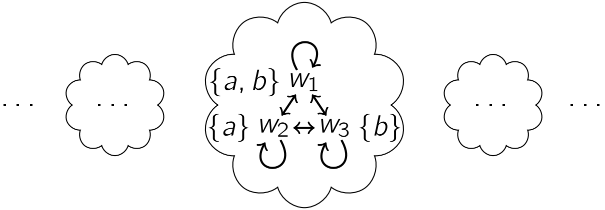
$\mathcal{F} \models \neg\Box\varphi \to \Box\neg\Box\varphi$ **верно для любой формулы** φ

$$\Leftrightarrow$$

отношение $\mapsto$ симметрично

Можете попробовать самостоятельно доказать эти утверждения по аналогии с утверждением про рефлексивность

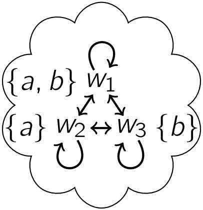
Следствие. **Модель Крипке идеального познающего субъекта — это модель, отношение переходов которой является отношением эквивалентности**



Пояснение

Пусть w_1 — мир достоверных фактов

Тогда класс эквивалентности w_1 состоит из всех миров, устройство которых не противоречит информации, которой располагает познающий субъект



Пояснение

- a — это факт, ...

$$\mathcal{I}, w_1 \models a$$

- ..., но моих знаний недостаточно, чтобы это утверждать, ...

$$\mathcal{I}, w_1 \not\models \Box a$$

- ..., но и опровергнуть a я тоже не могу,

$$\mathcal{I}, w_1 \models \Diamond a$$

- а что я точно знаю, так это то, что если не a, то обязательно b

$$\mathcal{I}, w_1 \models \Box(\neg a \rightarrow b)$$

В более широкой и «полезной» постановке задачи познающий субъект

- может изменять мир согласно своим скромным возможностям
- может взаимодействовать с другими такими же субъектами, обмениваясь с ними знаниями
- пытается достичь некоторой цели, кооперируясь или конкурируя с другими субъектами

Таких взаимодействующих субъектов принято называть агентами, а совокупность всех агентов с описанием их возможностей и целей — мультиагентной системой

Каждому агенту a такой системы присваивается своя эпистемическая модальность $\Box_a$: «агент a знает, что …»

Иногда рассматриваются и групповые модальности — например, $\Box_\forall$: «все агенты знают, что …»

Пример: задача о трёх мудрецах

Король призвал трёх мудрецов,
показал им три чёрные шапки и две белые, завязал глаза,
надел на мудрецов чёрные шапки, спрятал белые и развязал глаза

«Из пяти шапок, что я показал, три надеты на вас», — сказал король
«Знаете ли вы, какая на вас шапка?» — спросил король

«Нет, не знаю», хором ответили мудрецы

«Знаете ли вы, какая на вас шапка?» — повторил король

«Нет, не знаю», хором ответили мудрецы

«Знаете ли вы, какая на вас шапка?» — ещё раз повторил король

«**Да, чёрная**», хором ответили мудрецы

Как может выглядеть ход рассуждений мудрецов
в терминах эпистемической логики для мультиагентной системы?