

Математическая логика

mk.cs.msu.ru → Лекционные курсы → Математическая логика (318, 319/2, 241, 242)

Блок 44

Эпистемические логики

Лектор:

Подымов Владислав Васильевич

E-mail:

valdus@yandex.ru

Эпистемические логики

Эпистемическая логика (логика знаний) — это разновидность модальной логики, в которой модальность \square 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

$$\square\varphi \rightarrow \varphi \quad (\text{закон адекватности знания})$$

- ▶ мне известно, что именно я знаю

$$\square\varphi \rightarrow \square\square\varphi \quad (\text{закон позитивной интроспекции})$$

- ▶ мне известно, что именно я **не** знаю

$$\neg\square\varphi \rightarrow \square\neg\square\varphi \quad (\text{закон негативной интроспекции})$$

Эпистемические логики

Рассмотрим произвольную шкалу Крипке $\mathcal{F} = (W, R)$

Утверждение

$\mathcal{F} \models \Box\varphi \rightarrow \varphi$ верно для любой формулы φ
 \Leftrightarrow
отношение R рефлексивно

Утверждение

$\mathcal{F} \models \Box\varphi \rightarrow \Box\Box\varphi$ верно для любой формулы φ
 \Leftrightarrow
отношение R транзитивно

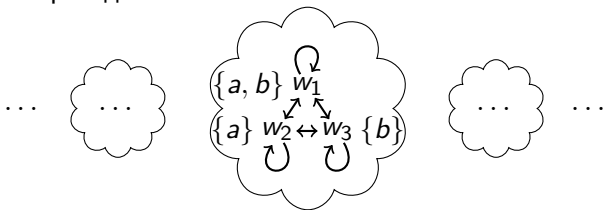
Утверждение. Пусть отношение R транзитивно. Тогда

$\mathcal{F} \models \neg\Box\varphi \rightarrow \Box\neg\Box\varphi$ верно для любой формулы φ
 \Leftrightarrow
отношение R симметрично

Доказательство. Попробуйте сами, если интересуетесь

Эпистемические логики

Следствие: модель Крипке идеального познающего субъекта — это модель \mathcal{I} , миры которой разбиваются на **классы эквивалентности** по отношению переходов:

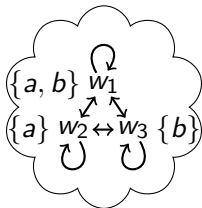


Пояснение

Пусть w_1 — мир достоверных фактов

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

Эпистемические логики



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

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

Эпистемические логики

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_V : «все агенты знают, что ...»

Эпистемические логики

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?