Имплицитное знание группы: трудности определения

В.В.Долгоруков

Международная лаборатория логики, лингвистики и формальной философии

НИУ ВШЭ

НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ

uAnalytiCon-2021

Екатеринбург, 14 мая 2021 г.

Имплицитное знание группы: трудности определения

Имплицитное знание группы: трудности определения

Идея имплицитного знания

Имплицитное групповое (групповое, дистрибутивное) знание – знание, которое есть у группы имплицитно: то есть, та информация, которую можно было бы извлечь, если бы агенты обменялись имеющимися у них сведениями.

Имплицитное знание группы: трудности определения

Идея имплицитного знания

Имплицитное групповое (групповое, дистрибутивное) знание – знание, которое есть у группы имплицитно: то есть, та информация, которую можно было бы извлечь, если бы агенты обменялись имеющимися у них сведениями.
История идеи : Я. Хинтикка, Р. Ауманн, Ф. Хайек и др.

Пример

- a знает, что p, но не гарантируется, что он осведомлен о $p \to q$
- b знает, что $p \to q$, но не гарантируется, что он осведомлен о p
- группа ab имплицитно знает, что q

Формализация интуиции

Определение

Моделью Крипке называется $\mathcal{M} = (W, \{\sim_i\}_{i \in \mathcal{A}}, V)$, где $\mathcal{A}$ - конечное множество агентов и

- $W \neq \emptyset$ – множество возможных миров
- $\sim_i \subseteq W \times W$ – отношение достижимости для агента i (в это докладе предполагается, что $\sim_i$ отношение эквивалентности)
- $V : Var \mapsto \mathcal{P}(W)$ – функции оценки

Имплицитное знание: семантика

Определение

Модальный оператор $K_i\varphi$ читается как «агент i знает, что φ».

$$\mathcal{M}, w \models K_i\varphi \text{ е.т.е. } \forall w'(w \sim_i w' \Rightarrow \mathcal{M}, w' \models \varphi)$$

Имплицитное знание: семантика

Определение [Van Der Hoek, van Linder, Meyer 1999] (≈ «group knowledge»)

Модальный оператор $I_G\varphi$ читается как «группа агентов G имплицитно знает, что φ».

$$\mathcal{M}, w \models I_G\varphi \text{ е.т.е. } \forall i \in G\ \exists \varphi_i \in \mathcal{L}_K :$$
$$\text{(a) } \mathcal{M}, w \models \bigwedge_{i \in G} K_i\varphi_i \text{ и (b) } \models \bigwedge_{i \in G} \varphi_i \to \varphi$$

Дистрибутивное знание: семантика

Определение [Halpern, Moses 1985]

$$\mathcal{M}, w \models D_G\varphi \iff \forall w'(w(\bigcap_{i\in G} \sim_i)w' \Rightarrow \mathcal{M}, w' \models \varphi)$$

Другие виды группового знания

Другие операторы:

- $\mathcal{M}, w \models S_G\varphi \iff \mathcal{M}, w \models \bigvee_{i \in G} K_i\varphi$
- $\mathcal{M}, w \models E_G\varphi \iff \mathcal{M}, w \models \bigwedge_{i \in G} K_i\varphi$ или
- $\mathcal{M}, w \models E_G\varphi \iff \forall w'(w(\bigcup_{i \in G} \sim_i)w' \Rightarrow \mathcal{M}, w' \models \varphi)$
- $\mathcal{M}, w \models C_G\varphi \iff \forall w'(w(\bigcup_{i \in G} \sim_i)^* w' \Rightarrow \mathcal{M}, w' \models \varphi)$, где R^* - рефлексивное транзитивное замыкание отношения R.
- $C_G\varphi = E_G\varphi \wedge E_G E_G\varphi \wedge \ldots$

$D_G\varphi$ vs $S_G\varphi$

$D_G\varphi$ vs $S_G\varphi$

Я не могу сделать горшок из глины, но человечество может.

$D_G\varphi$ vs $S_G\varphi$

Я не могу сделать горшок из глины, но человечество может.

vs.

$D_G\varphi$ vs $S_G\varphi$

Я не могу сделать горшок из глины, но человечество может.

vs.

Я не могу сделать автомобиль, но человечество может.

$D_G\varphi$ vs $S_G\varphi$

Я не могу сделать горшок из глины, но человечество может.
vs.
Я не могу сделать автомобиль, но человечество может.

D-технологии vs. S-технологии

- D-технологии: автомобиль, телефон и т.д.
- S-технологии: «первобытные» технологии, ремесла и т.д.

Как соотносятся I_G и D_G?

Как соотносятся I_G и D_G?

Имплицитное знание $\neq$ дистрибутивное знание, но

Как соотносятся I_G и D_G?

Имплицитное знание $\neq$ дистрибутивное знание, но различие между двумя визами знания не эпистемологическое, а «техническое» (связано с чувствительностью оператора к «лишним» мирам в модели).

Оператор D_G не выразим в базовом модальном языке

Как доказать невыразимость?

Оператор D_G не выразим в базовом модальном языке

Как доказать невыразимость?
Бисимуляция моделей влечет их модальную эквивалентность
(в базовом языке).

Понятие бисимуляции

Определение

Пусть $\mathcal{M} = (W, R, V)$ и $\mathcal{M}' = (W', R', V')$ – модели Крипке, тогда отношение $Z \subseteq W \times W'$ называется **бисимуляцией**, если выполняются следующие условия:

$$\begin{aligned} &\text{(var)} \quad xZx' \Rightarrow \mathcal{M}, x \models p \Leftrightarrow \mathcal{M}', x' \models p \\ &\text{(zig)} \quad (xRy \wedge xZx') \Rightarrow \exists y'(x'R'y' \wedge yZy') \\ &\text{(zag)} \quad (x'R'y' \wedge xZx') \Rightarrow \exists y(xRy \wedge yZy') \end{aligned}$$

Запись «$Z : (M, w) \simeq (M', w')$» означает, что Z является бисимуляцией, связывающей точки w и w'. Запись «$(M, w) \simeq (M', w')$» означает, что модели бисимулируют, то есть, существует отношение Z, являющееся бисимуляцией.

Теорема о бисимуляции

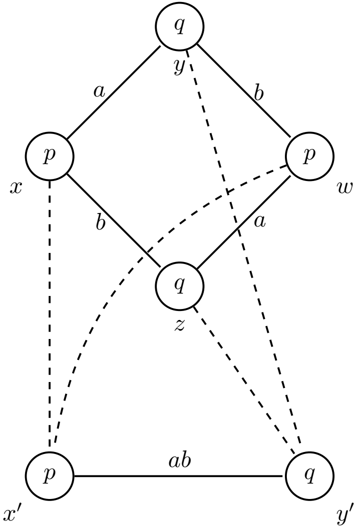
Если $(M, x) \simeq (M', x')$, то $(M, x) \equiv_{ML} (M', x')$. То есть, если две модели бисимулируют, то они модально эквивалентны.

Для конечных моделей (еще точнее – для моделей с конечным ветвлением) верна и обратная импликация.

1	$(M, x) \simeq (M', x')$	
2	$M, x \models \Box\varphi$	$\rhd : M', x' \models \Box\varphi$
3	$\boxed{y'_o}\ x'R'y'_0$	$\rhd : M', y'_0 \models \varphi$
4	xZx'	1
5	$\exists y(xRy \wedge yZy'_0)$	2, 3 (zag)
6	$\boxed{y_0}\ xRy_0 \wedge yZy'_0$	$\mathcal{M}', y'_0 \models \varphi$
7	xRy_0	5 $E_\wedge$
8	yZy'_0	5 $E_\wedge$
9	$\mathcal{M}, y_0 \models \varphi$	2, 7
10	$\mathcal{M}', y'_0 \models \varphi$	8, 9 по п.и.
11	$\mathcal{M}', y'_0 \models \varphi_1$	5, 6–10 $E_\exists$
12	$\forall y'(x'R'y' \Rightarrow \mathcal{M}', y' \models \varphi)$	3–11 $I_\forall$
13	$M', x' \models \Box\varphi_1$	12 по опр. $\Box$
14	$M, x \models \Box\varphi \Rightarrow M', x' \models \Box\varphi$	2–13 $I_\Rightarrow$

($\Leftarrow$) Аналогично, нужно использовать (zig).

$\mathcal{M}, x \models D_{ab}p$, но $\mathcal{M}', x' \not\models D_{ab}p$, хотя $(\mathcal{M}, x) \equiv_{ML} (\mathcal{M}', x')$



Оператор I_G безопасен для бисимуляции (bisimilar safe)

Почему?

- I_G определяется через K_i и логическое следование
- K_i – элемент базового языка (нужно добавить индексы в zig и zag для мультимодального языка), для которого работает теорема о бисимуляции
- логическое следование не зависит от конкретной модели и мира

Но все не так в классе конечных модельно различимых моделей [Van Der Hoek, van Linder, Meyer 1999]

Определение

Будем говорить, что формула φ **различает** модели $(\mathcal{M}, w)$ и $(\mathcal{M}', w')$, если $\mathcal{M}, w \models \varphi$ и $\mathcal{M}', w' \not\models \varphi$.

Определение

Будем называть модель Крипке **модально различимой** если любые две различные точки этом модели различимы какой-либо формулой.

Ключевая идея

Ключевая идея

1. $\models I_G\varphi \to D_G\varphi$

Ключевая идея

1. $\models I_G\varphi \to D_G\varphi$
2. Раз мы имеем дело с модально различимой моделью, то для каждого мира w_i найдется его характеристическая формула χ_i.

Ключевая идея

1. $\models I_G\varphi \to D_G\varphi$
2. Раз мы имеем дело с модально различимой моделью, то для каждого мира w_i найдется его характеристическая формула χ_i.
3. Пусть для $k \in G$ формула φ_k строится следующим образом:

$$\varphi_k := \bigvee_{i \in [w]_k} (\chi_i \wedge \bigwedge_{\substack{j \neq i, \\ j \in W}} \neg\chi_j)$$

$[w]_k := \{w' \in W \mid w \sim_k w'\}$

Ключевая идея

1. $\models I_G\varphi \to D_G\varphi$
2. Раз мы имеем дело с модально различимой моделью, то для каждого мира w_i найдется его характеристическая формула χ_i.
3. Пусть для $k \in G$ формула φ_k строится следующим образом:
$$\varphi_k := \bigvee_{i \in [w]_k} (\chi_i \wedge \bigwedge_{\substack{j \neq i, \\ j \in W}} \neg\chi_j)$$
$[w]_k := \{w' \in W \mid w \sim_k w'\}$
4. условие (a) выполняется, поскольку $[[\varphi_k]]_{\mathcal{M}} = [w]_k$
$[[\varphi]]_{\mathcal{M}} := \{w \in W \mid \mathcal{M}, w \models \varphi\}$

Ключевая идея

1. $\models I_G\varphi \to D_G\varphi$
2. Раз мы имеем дело с модально различимой моделью, то для каждого мира w_i найдется его характеристическая формула χ_i.
3. Пусть для $k \in G$ формула φ_k строится следующим образом:
$$\varphi_k := \bigvee_{i \in [w]_k} (\chi_i \wedge \bigwedge_{\substack{j \neq i, \\ j \in W}} \neg\chi_j)$$
$[w]_k := \{w' \in W \mid w \sim_k w'\}$
4. условие (a) выполняется, поскольку $[[\varphi_k]]_{\mathcal{M}} = [w]_k$
$[[\varphi]]_{\mathcal{M}} := \{w \in W \mid \mathcal{M}, w \models \varphi\}$
5. условие (b) выполняется в силу «резолюции»:
$$A \vee B, C \vee D, \neg(B \wedge C) \models A \vee D$$

Модификация условия извлечения информации

$\mathcal{M}, w \models I_G\varphi$ е.т.е.$\forall i \in G\ \exists \varphi_i \in \mathcal{L}_K$:

$$\text{(a) } \mathcal{M}, w \models \bigwedge_{i \in G} K_i\varphi_i$$

$$\text{(b) } \models \varphi_G \to \varphi, \text{ где } \varphi_G := \bigwedge_{i \in G} \varphi_i.$$

5 оттенков имплицитного знания

Варианты условия (b):

$$\models \varphi_G \to \varphi$$

5 оттенков имплицитного знания

Варианты условия (b):

$$\models \varphi_G \to \varphi$$

1. $\mathcal{M}, w \models [\forall](\varphi_G \to \varphi)$

5 оттенков имплицитного знания

Варианты условия (b):

$$\models \varphi_G \to \varphi$$

1. $\mathcal{M}, w \models [\forall](\varphi_G \to \varphi)$
2. $\mathcal{M}, w \models C_G(\varphi_G \to \varphi)$

5 оттенков имплицитного знания

Варианты условия (b):

$$\models \varphi_G \to \varphi$$

1. $\mathcal{M}, w \models [\forall](\varphi_G \to \varphi)$
2. $\mathcal{M}, w \models C_G(\varphi_G \to \varphi)$
3. $\mathcal{M}, w \models E_G(\varphi_G \to \varphi)$

5 оттенков имплицитного знания

Варианты условия (b):

$$\models \varphi_G \to \varphi$$

1. $\mathcal{M}, w \models [\forall](\varphi_G \to \varphi)$
2. $\mathcal{M}, w \models C_G(\varphi_G \to \varphi)$
3. $\mathcal{M}, w \models E_G(\varphi_G \to \varphi)$
4. $\mathcal{M}, w \models S_G(\varphi_G \to \varphi)$

5 оттенков имплицитного знания

Варианты условия (b):

$$\models \varphi_G \to \varphi$$

1. $\mathcal{M}, w \models [\forall](\varphi_G \to \varphi)$
2. $\mathcal{M}, w \models C_G(\varphi_G \to \varphi)$
3. $\mathcal{M}, w \models E_G(\varphi_G \to \varphi)$
4. $\mathcal{M}, w \models S_G(\varphi_G \to \varphi)$
5. $\mathcal{M}, w \models \varphi_G \to \varphi$

5 оттенков имплицитного знания

Варианты условия (b):

$$\models \varphi_G \to \varphi$$

1. $\mathcal{M}, w \models [\forall](\varphi_G \to \varphi)$
2. $\mathcal{M}, w \models C_G(\varphi_G \to \varphi)$
3. $\mathcal{M}, w \models E_G(\varphi_G \to \varphi)$
4. $\mathcal{M}, w \models S_G(\varphi_G \to \varphi)$
5. $\mathcal{M}, w \models \varphi_G \to \varphi$

5 оттенков имплицитного знания

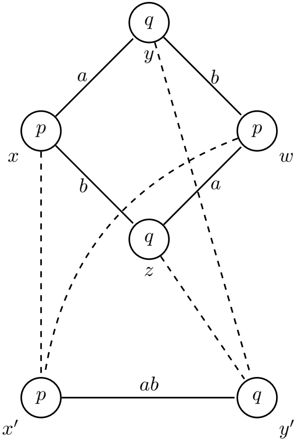
Варианты условия (b):

$$\models \varphi_G \to \varphi$$

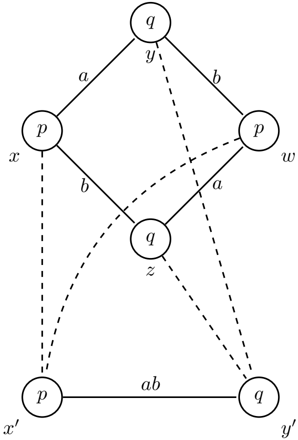
1. $\mathcal{M}, w \models [\forall](\varphi_G \to \varphi)$
2. $\mathcal{M}, w \models C_G(\varphi_G \to \varphi)$
3. $\mathcal{M}, w \models E_G(\varphi_G \to \varphi)$
4. $\mathcal{M}, w \models S_G(\varphi_G \to \varphi)$
5. $\mathcal{M}, w \models \varphi_G \to \varphi$

Оказывается, что в классе конечных модально различимых моделей достаточно уже условия (4).

В чем разница между к.м.р. моделями и не-к.м.р?



В чем разница между к.м.р. моделями и не-к.м.р?



Сверху и снизу одна и та же модель? Какие ситуации она различает? Допустимы ли в модели копии одного и того же мира?

Открытые вопросы

Какими свойствами будет обладать имплицитное знание в расширенном языке (с модальностью D_G)?

Слабое имплицтиное знание

$\mathcal{M}, w \models I^*_G\varphi$ е.т.е. $\forall i \in G\ \exists \varphi_i \in \mathcal{L}_{KD}$:
(a) $\mathcal{M}, w \models \bigwedge\limits_{i \in G} K_i\varphi_i$ и
(b) $\mathcal{M}, w \models D_G\,(\bigwedge\limits_{i \in G} \varphi_i \to \varphi)$

Литература

1. Fagin R., Halpern J.Y., Vardi M.Y. What can machines know? // Journal of the ACM. 1992. Vol. 39, № 2. P. 328–376.
2. Halpern J.Y., Moses Y., A guide to the modal logics of knowledge and belief // Proceedings of IJCAI-85. Los Angeles, CA, 1985. P. 480–490
3. Van Der Hoek W., Van Linder B., Meyer J.J. Group knowledge is not always distributed (neither is it always implicit) // Mathematical social sciences. 1999. Vol. 38, № 2. P. 215–240.
4. Van der Hoek W., Meyer J.J. A Complete Epistemic Logic for Multiple Agents. 1997 . P. 35–68.
5. Wáng Y.N., Ågotnes T. Public announcement logic with distributed knowledge: Expressivity, completeness and complexity // Synthese. 2013. Vol. 190. P. 135–162.