Информированность агентов о чужой информированности: формализация и вычисление

Денис Федянин
dfedyanin@{inbox, hse, ipu, nes, mipt}.ru

В рамках проекта РНФ «Логико-когнитивные модели рассуждений: принципы демаркации нормативного и дескриптивного» No 23-18-00695

Мотивирующий пример

		Такты								
		1	2	3	4	5	6	7	8	9
Агенты	1	1	1	1	1	0	0	1	1	0
	2	0	1	0	1	1	1	1	1	0
	3	0	1	0	0	0	1	0	1	0
	4	1	1	0	1	0	1	0	1	1
События	a=0	b=1	c=a+b	b=4	d=1	b=a*c	e=5	e=d+1	c=2	a=0
	b=0			c=3	b=2	c=d+1	c=b	d=b+2	b=1	b=2
	c=0			a=1						
	d=0									
	e=0									

Рис.: Таблица присутствия агентов

Вопросы:

- Чему равно по мнению агента 3 значение переменной b после 9-го такта? а агента 2? а агента 4?
- Какое значение переменной c по мнению агента 3 агент 2 считает наиболее вероятным после 9-го такта?

arXiv:2506.09250v1 [cs.AI] 10 Jun 2025

The Illusion of the Illusion of Thinking A Comment on Shojaee et al. (2025)

C. Opus
Anthropic

A. Lawsen
Open Philanthropy

(June 10, 2025)

Abstract

Shojaee et al. (2025) report that Large Reasoning Models (LRMs) exhibit ”accuracy collapse” on planning puzzles beyond certain complexity thresholds. We demonstrate that their findings primarily reflect experimental design limitations rather than fundamental reasoning failures. Our analysis reveals three critical issues: (1) Tower of Hanoi experiments systematically exceed model output token limits at reported failure points, with models explicitly acknowledging these constraints in their outputs; (2) The authors' automated evaluation framework fails to distinguish between reasoning failures and practical constraints, leading to misclassification of model capabilities; (3) Most concerningly, their River Crossing benchmarks include mathematically impossible instances for $N \geq 6$ due to insufficient boat capacity, yet models are scored as failures for not solving these unsolvable problems. When we control for these experimental artifacts, by requesting generating functions instead of exhaustive move lists, preliminary experiments across multiple models indicate high accuracy on Tower of Hanoi instances previously reported as complete failures. These findings highlight the importance of careful experimental design when evaluating AI reasoning capabilities.

Убеждения агентов

Обозначим $B_i(x = x_0)$ - логическое утверждение, что агент i верит, что значение переменной (наиболее вероятное) x это x_0.
И сразу

$$\theta_i = \{\theta \mid B_i(x = \theta)\}.$$

Аналогично обозначим $B_j B_i(x = x_0)$ - логическое утверждение, что агент j верит, что агент i верит, что значение переменной (наиболее вероятное) x это x_0.

$$\theta_{ij} = \{\theta \mid B_i B_j(x = \theta)\}.$$

Как же их посчитать?

Возможные предположения

1. Агент считает, что в его отсутствие система не меняла своего состояния (когнитивная инерция)
2. Агент, считает, что в его отсутствие система меняла свое состояние. Но тогда ему нужна модель поведения системы. Он может попровать использовать метод наибольшего правдоподобия чтобы оценить параметры состояния системы, если считает, что они случайны (например, нормальны) или построить тренд, но здесь нужно еще больше предположений...
3. Агент может сохранять неопределенность и считать, что возможны несколько вариантов состояний системы. Но тогда вопрос - а какие ему считать возможными? Кроме того их может быть бесконечно много, а с этим иногда трудно работать...
4. ...

Остановимся пока на первом предположении.

Динамика состояния системы

Пусть начальное состояние системы s_0. Обозначим $e(s)$ - состояние системы, находившейся в состоянии s после события e, мы также можем записать динамику изменения системы с помощью производящей функции:

$$f = \sum_i e_i z^i,$$

где e_i - преобразование системы событием e_i. И конечное состояние, как

$$s_F = e(f, s_0) = e\left(\sum_i e_i z^i, s_0\right).$$

Обозначим $a_{it} = 1$ присутствие агента i на такте t.
И через $a_{it} = 0$ его отсутствие.

Вычисления убеждений

Тогда

$$B_i\left(s^F = e\left(\sum_{m=1}^{k}(e_m z^m)^{a_{im}}, s_0\right)\right)$$

Также можно построить и убеждения более высокого порядка

$$B_{i_1}\ldots B_{i_q}\left(s^F = e\left(\sum_{m=1}^{k}(e_m z^m)^{\prod_{r=1}^{q} a_{i_r m}}, s_0\right)\right)$$

или

$$\theta_{i_1\ldots i_m} = e\left(\sum_{m=1}^{k}(e_m z^m)^{\prod_{r=1}^{q} a_{i_r m}}, s_0\right)$$

Свойства

- Коммутативность

$$B_i B_j(x = x_0) = B_j B_i(x = x_0)$$

- Идемпотентность

$$B_i B_i(x = x_0) = B_i(x = x_0)$$

Для любой формулы $B_1 \dots B_l(x = x_0)$ найдется эквивалентная формула длиной не более количества агентов. Это полезно для вычисления общего знания.

Нормативность и дескриптивность

- Кратко и не вдаваясь в подробности проиллюстрируем различие дескриптивного и нормативного (ожидаемого поведения) агентов.
- Выделим пару агентов с индексами 1 и 2, и первого из них назначим ведущим, а второго - ведомым и положим, что мнение ведущего - нормативное.
- Тогда мы можем предсказать что будет думать ведоомый (в рамках модели) о том, насколько его реальное понимание отличается от того, которое по его мнению имеет его начальник (и которое по мнению ведомого является для него нормативным)

$$L_2 = (\theta_2 - \theta_{21})^2 \to \min$$

- В тоже время можно измерить и то, насколько будет недоволен точностью убеждений своего подчиненного ведущий

$$L_1 = (\theta_1 - \theta_{12})^2 \to \min$$

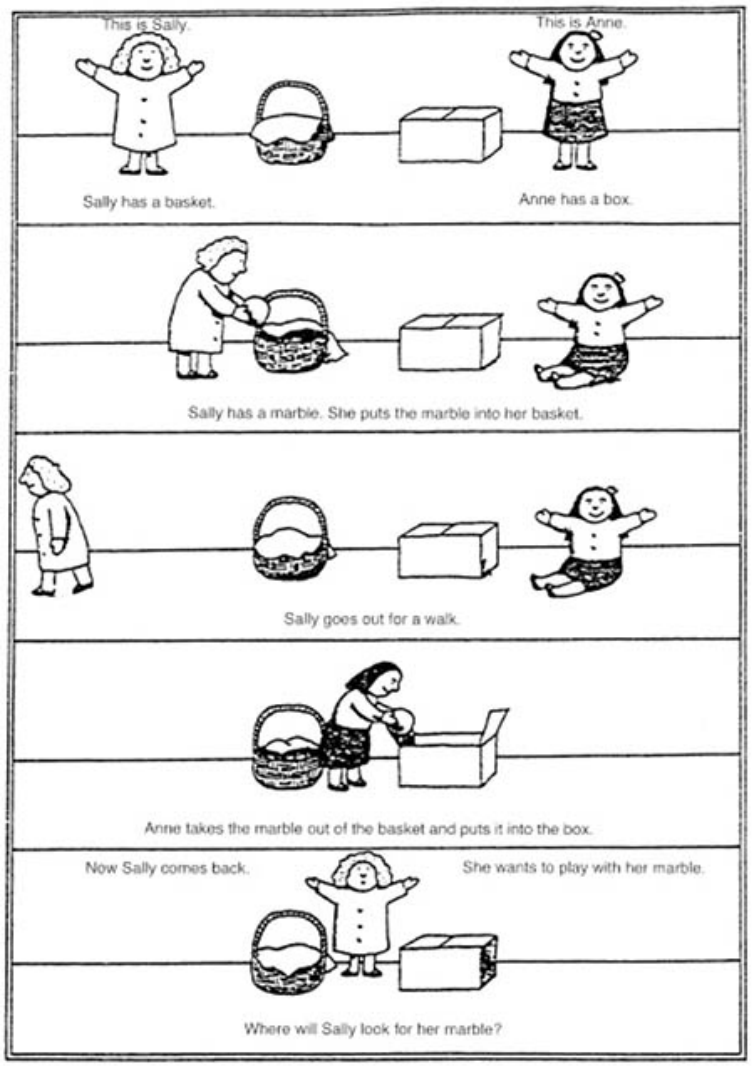
Обобщения

- По способу построения видно, что дискретность количества событий непринципиальна - их может быть континуум (если мы говорим про теоретические исследования) - например измеряемые значения скорости (или ускорения) во времени. Надо только найти времена, когда интересующие нас агенты были вместе и найти интеграл в этой области - получим ожидаемую координату (кстати и агентов может быть бесконечное количество, если задано, как мы ищем пересечение их времен присутствия).
- Также не приниципиально и то, что агенты считают что система в их отсутствие не менялась - достаточно просто аккуратно прописать их предположения как она изменялась в их отсутствие (но важно, чтобы все остальные тоже их знали)

Обобщения

- Если мы не уверены, что агенты внимательны (например они в телефонах), то такая схема не пройдет, однако можно перейти к нечеткой логике (нечетким множествам), где $\mu_i(t) = a_{it} \leq 1$ будет задавать степень уверенности, что агент i внимателен в момент t. А увереннность в том, что агенты были в момент t будут вычисляться как минимум от уверенностей в их индивидуальном присутствии (т.е. как раз соответствует одному из определений конъюнкции для многозначной логики). Впрочем, в этом случае расчет именно убеждений гораздо более сложен, так как у агентов сохраняется неопределенность.

Салли, Энн и связь с модальной логикой



Ann	*Sally*	
1	1	$a = 0; b = 0$
1	1	$a = 1; b = 0$
1	0	$a = 0; b = 1$
1	1	$\emptyset$

DEL-based Epistemic Planning for Human-Robot Collaboration: Theory and Implementation

Thomas Bolander, **Lasse Dissing**, **Nicolai Herrmann**
Technical University of Denmark
{tobo, ldiha}@dtu.dk, nicolai@herrmann.dk

Proceedings of the 18th International Conference on Principles of Knowledge Representation and Reasoning
Main Track

Algorithm 2: *Plan*

Input: A planning task $\Pi = (s_0, A, \gamma)$ for agent $i \in \mathcal{A}$
Output: Policy π for Π, or *failure*

```
initialize(i, s_0)
while not frontier.empty() do
    s ← frontier.dequeue()
    for each j:a ∈ A do
        if not s^j.applicable(j:a) then continue
        s' ← s^j ⊗ j:a
        s' ← OrderedPartitionRefinement(s')
        if s' ∈ S_and then continue
        S_and ← S_and ∪ {s'}
        for each s'' ∈ Globals(s') do
            s'' ← OrderedPartitionRefinement(s'')
            if s'' ∈ S_or then continue
            S_or ← S_or ∪ {s''}
            if s'' ⊨ γ then
                update_solved_dead(s'')
            else
                frontier.enqueue(s'')
    update_solved_dead(s)
    if solved(root) then return extract_policy()
    if dead(root) then return failure
return failure
```

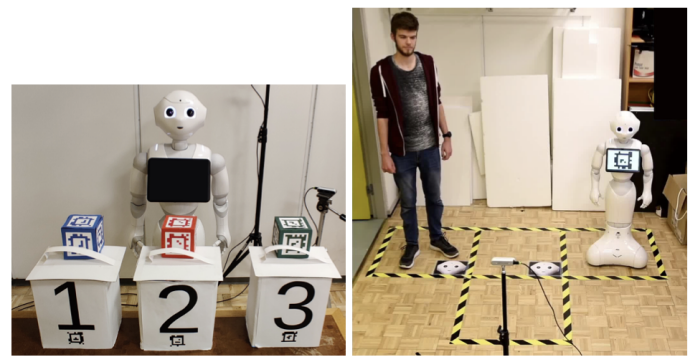


Figure 5: The Pepper robot used in our experiments. On the left we see part of the setup for the cubes and boxes domain. On the right we see a MAPF/DU problem instance where the barricade tape marks the grid.

bottom-up iterative traversal, where we start by marking all solved leaf nodes with a cost of 0 and all other nodes with a special undefined cost. Then in the i'th iteration, we perform the following check for all parents of nodes with cost $i-1$: if the parent has an undefined cost and is an or-node, we mark it with cost i; if the parent is an and-node and all its children have defined costs, we mark it with the maximal cost of its children. We continue this until the root node is reached. The second traversal is top-down, starting at the

Связь с модальной логикой

$$M = (W, \{R_i\}_{i \in A}, V)$$

$$W \neq \varnothing$$

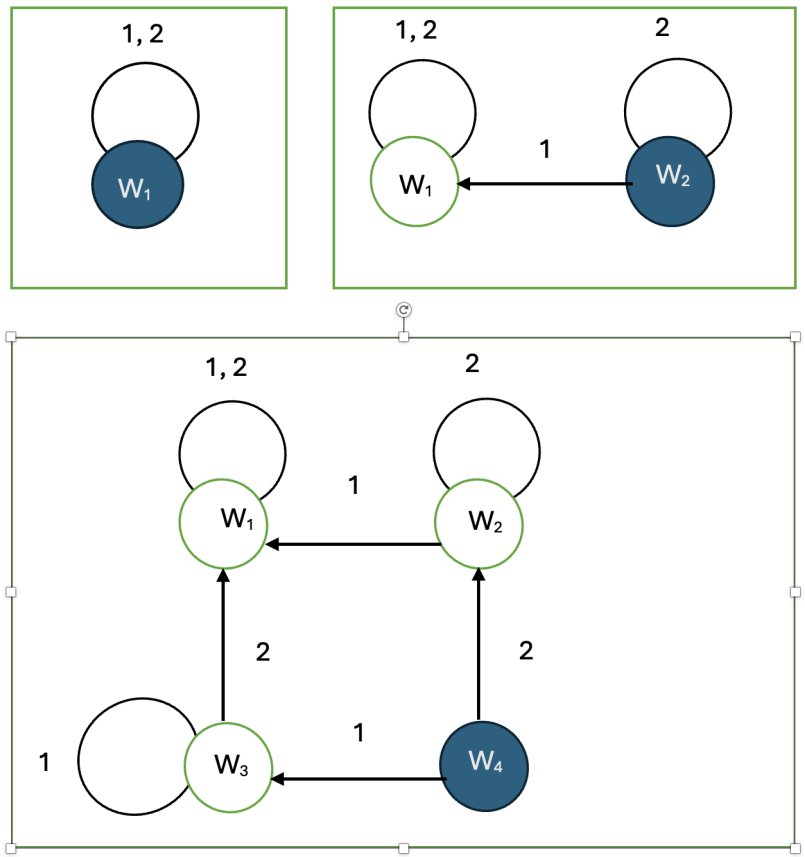
$$R_i \in W \times W$$

$$V : P \to 2^W$$

$$\varphi := p \mid \neg\varphi \mid \varphi \vee \varphi \mid B_i\varphi$$

$$M, w \models p \iff w \in V(p)$$

$$M, w \models B_i\varphi \iff \forall x(wR_ix \to M, x \models \varphi)$$



Бонусный слайд

- Модель убеждений (эпистемические состояния) (и сами убеждения) можно использовать как дополнительные параметры состояния для построения более точной модели RL - ведь то, что знает соперник очень даже может влиять на то, что какая стратегия будет оптимальной.
- Особенно в играх на скрытность и многопользовательских - в Dota 2 например не всегда тривиален выбор куда и когда поставить варды и видел ли вашу команду соперник при применении ей дыма.

Спасибо за внимание!