标题 title

作者 *author*

2024 年 9 月 15 日

前言

目录

第一部分

AI 的逻辑

第二部分

信息与数据

第三部分

决策与优化

第四部分

逻辑与博弈

第五部分

认知逻辑

第十章　共同知识，Bayes 博弈，Aumann 知识算子

真实世界是一个巨大的游戏，参与者之间的信息不对称是一个普遍存在的现象，这些不对称往往形成了“丛林法则”。“丛林法则”从来不是一个书面的规则。刚来美国旅游的人可能会非常担心自己被抢劫；然而，“盗亦有道”，绝大部分时候，抢劫犯只会抢二十美元，以便吃一顿饭。如果坏了规矩，反而会被惩罚。

我们可以考虑一个看似非常疯狂的想法：既然有丛林法则，何不把“一次抢劫最多二十美元”写入法律，这会有什么区别呢？抛开道德和法律的问题，这样做似乎不无道理。

1982 年，经济学家 Alvin E. Roth 和 J. Keith Murnighan 对讨价还价这一经典的市场博弈现象进行了同样的实验。实验中，有两个玩家，他们可能会收到特定价值的奖品，每个玩家的奖品价值可以不同。然而玩家并不总是能获得奖品，他有一个概率获得这一奖品。他们需要在规定的时间内就他们各自获得奖品的概率进行讨价还价。

具体来说，玩家其实在商讨如何分配一张 100% 概率的“彩票券”，这个券决定了每个玩家赢得奖品的概率。例如，甲如果获得 40% 的彩票券，就有 40% 的概率赢得奖品，60% 的概率一无所获；而乙则完全相反，有 60% 的概率赢得奖品，40% 的概率一无所获。

然而，如果在规定时间内未能达成协议，那么所有玩家都将一无所获。因此，只有在双方就彩票券的分配达成了协议，并且该玩家在随后的抽奖中中奖时，玩家才能获得相应的奖品。否则，他将得不到任何奖励。我们将这种每个玩家只有两种可能金钱收益的游戏称为“二元彩票游戏”。

在实际的实验中，玩家甲的奖品价值 20 美元，玩家乙的奖品价值为 5 美元，谈判时间限制为 12 分钟。在实验开始前，实验人员会公开告知甲乙完整的游戏规则，包括奖品价值和谈判时间限制。实验人员还会公开告知甲乙关于他们私有信息的情况，例如，

> “在游戏开始后，你们奖品的价值会被告诉对方。”
> “在游戏开始后，甲的奖品价值会被告诉乙，但乙的不会被告诉甲。”

实验人员也可以选择不告知这些信息。

实验开始后，实验人员会告知甲乙他们各自的奖品价值，并且选择性地告知甲（或乙）乙（或甲）的奖品价值，这些信息要和实验开始前的信息保持一致。然后，甲乙开始商讨如何分配彩票券。如果在规定时间内达成了协议，那么他们将按照协议分配彩票券。如果未能达成协议，那么他们将一无所获。

上面的实验设置中，在游戏开始前是否公开宣告双方的信息结构，和上面那个疯狂的想法是一样的。然而，在讨价还价的情景下，我们似乎不会觉得这有什么区别。然而，实验结果却显示，在信息不对称的情况下，双方信息结构的公开宣告会对博弈结果产生显著的影响。

谈判过程展现出了明显的策略性行为。例如，甲（20 美元的玩家）通常不会提及自己奖品的价值；但如果游戏开始前宣布过“乙（5 美元的玩家）不清楚甲的奖品价值”，20 美元的玩家往往会虚报自己的奖品。一个典型的例子是：“我知道你的奖品是 5 美元。我的只有 2 美元。所以我应该得到超过 50% 的份额。”另一方面，当乙知道甲的奖品价值时，往往会透露信息。这两种策略通常都不被对方相信。

更重要的是，实验结果显示，当实验人员在游戏开始前宣布双方的信息结构时，玩家变得更加没有策略性，博弈的结果很可能没有达到 Nash 均衡：当只有 20 美元的玩家知道两个奖品时，其总体平均收益（包括达成协议和未达成协议的情况）为 34.9，显著低于 5 美元玩家的相应收益 53.6。

另一方面，当双方的信息结构在游戏开始前不被公开宣告时，玩家表现出更多的策略性。在这种情况下，因为他们不能确定对手是否不知道自己的奖品，所以玩家无法像前一种情况那样自由地谎报自己的奖品价值。然而，因为玩家都不知道对方是否指导双方奖品的价值，所以，如果一个玩家知道两个奖品的价值，他完全可以假装自己只知道自己的奖品价值。由于更复杂的策略性行为，博弈的结果往往达到 Nash 均衡！

上面的故事告诉我们，信息结构（也就是我们*知道*什么）是否被公开宣告，对于博弈的结果有着重要的影响。实际上，这一问题是一个关于*共同知识*的问题。

在本章，我们将更系统地讨论共同知识，并定义 *Bayes 博弈*，它是我们研究博弈论中信息结构的基本语言。最后，我们将介绍 *Aumann 知识算子*，它是从 Bayes 博弈中起源的，一个关于“知识”的数学模型。

§10.1 “泥泞的孩童”谜题

我们先从一个经典的谜题开始。

有 n 个孩子在玩泥巴，他们互相泼泥巴. 母亲告诉孩子们，如果他们脸上沾上了泥

巴，会受到严厉的惩罚. 孩子们不能看到自己的脸，但是可以看到其他所有人的脸. 所有孩子都希望保持自己的脸干净，但是弄脏别人的脸. 此时，孩子的父亲出现了，于是，孩子们停止泼泥巴. 孩子们互相不说话. 父亲看到了 k（$k \geq 1$）个人脸上有泥巴，于是宣布："你们至少有一个人脸上沾了泥巴."之后，父亲会公开地问若干轮如下问题："你们知道自己脸上有泥巴了吗？"孩子们回答"知道"或者"不知道". 假设孩子们观察力敏锐、聪慧且诚实，并且每一轮他们都同时回答. 接下来会发生什么？

假设有 k 个孩子脸上有泥巴. 谜底：在前 $k-1$ 轮中，所有孩子都会说"不知道"，在第 k 轮中，所有脸上有泥巴的孩子都会说"知道". 这一结论的论证来源于对 k 的归纳.

当 $k=1$ 时，脸上沾满泥巴的孩子看到其他人都没有泥巴. 既然他知道至少有一个孩子的脸上有泥巴，他就能推出那个人肯定是他自己.

现在假设 $k=2$，脸上沾满泥巴的孩子是 a 和 b. 一开始，因为他们分别看到了对方的脸上有泥巴，所以他们每个人都回答"不知道". 但是，当 b 回答"不知道"时，a 意识到他自己肯定是脸上有泥巴的那个孩子，否则 b 就会在第一轮中知道泥巴在他的脸上，并回答"知道". 因此，a 在第二轮回答"知道". b 也会通过同样的推理得出相同的结论.

现在假设 $k=3$，脸上沾满泥巴的孩子分别是 a，b 和 c. 孩子 a 的论证如下. 假设我没有泥巴落在脸上. 根据 $k=2$ 的情况，b 和 c 在第二轮都会回答"是". 他们没有这样做，我意识到假设是错误的，我的脸上也有泥巴. 因此在第三轮我会回答"知道". b 和 c 的论证也是类似的.

$k=3$ 的论证具有一般性，对一般的 k 也成立.

注. "泥泞的孩童"还有其他流行的陈述方式，比如"蓝眼睛红眼睛". 一个岛上有 100 个人，其中有 5 个红眼睛，95 个蓝眼睛. 这个岛有三个奇怪的宗教规则.

1. 他们不能照镜子，不能看自己眼睛的颜色.
2. 他们不能告诉别人对方的眼睛是什么颜色.
3. 一旦有人知道了自己的眼睛是红色，他就必须在当天夜里自杀.

岛民不知道具体有几个红眼睛.

某天，有个旅行者到了这个岛上. 由于不知道这里的规矩，所以他在和全岛人一起狂欢的时候，一不留神说了一句话："你们这里有红眼睛的人."假设这个岛上的人足够聪明，每个人都可以做出缜密的逻辑推理. 请问这个岛上将会发生什么？

那么，为什么会这样呢？如果 $k>1$，那么所有人都知道 p："至少有一个人脸上有泥巴". 那么父亲说这句话的意义是什么？如果父亲没有说 p，那么会发生什么？无论父亲问多少轮，所有孩子都只会回答"不知道"！（为什么）因此，父亲公开说了 p，这是谜题的关键.

假设 $k=2$，脸上沾满泥巴的孩子是 a 和 b. 在父亲宣布 p 之前，a 和 b 都知道 p. 然而，他们并不知道对方知道 p. a 可能会有两种想法：

- 我的脸上有泥巴，所以 b 知道 p.
- 我的脸上没有泥巴，b 是唯一一个有泥巴的，所以 b 不知道 p.

当父亲宣布 p 之后，a 知道了 b 知道 p. 当第一轮 b 回答"不知道"之后，a 可以用"b 知道 p"这一知识推出自己脸上有泥巴.

假设 $k=3$，脸上沾满泥巴的孩子是 a，b 和 c. 在父亲宣布 p 之前，a，b 和 c 不仅知道 p，而且知道彼此知道 p. 以 a 的视角看，b 能看到 c 脸上有泥巴，所以 a 知道 b 知道 p. 但是，a，b，c 都不知道所有人知道所有人知道 p！

用 $E^m p$ 表示所有人知道所有人知道……所有人知道（m 次）p. 在一般情况下，父亲没有宣布 p 之前，$E^k p$ 并不成立. 父亲宣布了 p 之后，对任意 $m \geq 1$，$E^m p$ 都成立！因此，父亲宣布 p 带来了*共同知识*. 有了共同知识，这一谜题就可以按照我们所讨论的方式进行下去.

我们曾经假设过所有人"观察力敏锐、聪慧且诚实". 然而，这一假设并不足够. 我们必须假设所有人都知道所有人"观察力敏锐、聪慧且诚实"，所有人都知道所有人都知道所有人"观察力敏锐、聪慧且诚实"，……换言之，我们需要假设"所有人观察力敏锐、聪慧且诚实"是共同知识. 假设还是只有两个孩子 a,b 脸上有泥巴. 假如 a 不知道 b 是诚实的，即便 b 回答了"不知道"，a 也无法从 b 的回答中得到任何额外的知识！

除了假设"所有人观察力敏锐、聪慧且诚实"是共同知识，我们还需要假设以下陈述是共同知识：

- 每个人都能看到所有除自己外的人.
- 每个人都听到了父亲说的话.
- 父亲是诚实的.
- 每个人都在每一轮进行了充分的推理.
- ……

任何假设的破坏都会导致之前的讨论失效. 那么，为什么父亲宣布 p 就可以让 p 变成共同知识呢？

所有人都*听到*父亲说 p 并不能产生共同知识. 假如父亲只是对每一个孩子单独宣布 p. 所有人并不知道所有人都知道 p，因而仅仅可以做到 Ep. 那么，所有人都*知道*所有人

听到父亲说 p 会如何呢？进一步假设每个孩子给每一个孩子都安装了窃听器，每个人都能够偷听每个人与父亲的谈话内容. 所有人并不知道所有人都知道所有人都知道 p，因而仅仅有 E^2p. 因此，父亲宣布 p 会产生共同知识的核心原因是*公开宣布*，此时对每一个 m 都有 $E^m p$.

"泥泞的孩童"谜题足以表明，关于"知道"的讨论远比想象的复杂. 关于"知道"和知识的研究在哲学中划归为*知识论*. 接下来，我们将使用模态逻辑来形式化关于"知道"和知识的讨论，这被称之为*认知逻辑*.

我们将介绍处理知识的两种数学模型：

- 一种源自 Aumann，Harsanyi 和 Rubinstein 等人，以 Bayes 概率论为基础，是偏经济学的学术风格；
- 另一种源自 Kripke，Hintikka 和 Halpern 等人，以模态逻辑为基础，是偏计算机科学和哲学的学术风格。

在这一章，我们主要讨论 Bayes 概率论的方法.

§10.2　不完全信息博弈（Bayes 博弈）

§10.3　Rubinstein 电子邮件博弈

§10.4　Aumann 知识算子

§10.5　习题

§10.6　章末注记

第六部分

附录：预备知识

参考文献

[Bre57] Leo Breiman. The Individual Ergodic Theorem of Information Theory. *The Annals of Mathematical Statistics*, 28(3):809–811, 1957.

[CT12] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. John Wiley & Sons, 2012.

[Huf52] David A. Huffman. A Method for the Construction of Minimum-Redundancy Codes. *Proceedings of the IRE*, 40(9):1098–1101, September 1952.

[Inf] Information | Etymology, origin and meaning of information by etymonline. https://www.etymonline.com/word/information.

[Jay02] Edwin T. Jaynes. *Probability Theory: The Logic of Science*. Cambridge University Press, 2002.

[KL51] S. Kullback and R. A. Leibler. On Information and Sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, 1951.

[LLG+19] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, and Luke Zettlemoyer. BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension, October 2019.

[McM53] Brockway McMillan. The Basic Theorems of Information Theory. *The Annals of Mathematical Statistics*, 24(2):196–219, June 1953.

[RHW86] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning internal representations by error propagation. In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1: Foundations*, pages 318–362. MIT Press, Cambridge, MA, USA, January 1986.

[Rob49] Robert M. Fano. *The Transmission of Information*. March 1949.

[Sha48] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, July 1948.

[Shi96] A. N. Shiryaev. *Probability*, volume 95 of *Graduate Texts in Mathematics*. Springer, New York, NY, 1996.

[Tin62] Hu Kuo Ting. On the Amount of Information. *Theory of Probability & Its Applications*, 7(4):439–447, January 1962.

[Uff22] Jos Uffink. Boltzmann's Work in Statistical Physics. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, summer 2022 edition, 2022.

[李 10] 李贤平. *概率论基础*. 高等教育出版社, 2010.

索引