

# 当博弈论遇上人工智能

计算机学院

余皓然

2021/5/29

# 主要内容

---



## 背景介绍

- ▶ 什么是博弈论
- ▶ 什么是机制设计
- ▶ 完全理性假设



## 与AI的结合

- ▶ 博弈论 + AI
- ▶ 机制设计 + AI

# 背景介绍

什么是博弈论

什么是机制设计

完全理性假设

# 什么是博弈论

## 精通博弈论有什么好处？

圆桌收录 日常经济学 · 博弈人生 >

题主不是一个功利主义者，只是在翻书的时候书中有很多涉及博弈论上的内容难以理解，起了兴趣，故有此一问 本题已加入知乎圆桌 » 日常经济学 · 博弈人生 ...[显示全部](#) ▾

精通博弈论的人，一般也不大愿意让别人知道自己精通博弈论。因为那样带来的麻烦，从长期看，可能会远多于赢得这种名声的好处。所以，题主应该看不到心中期望的答案啦。

整个二十四史，不都在反复证明一个道理——聪明又不懂得保护（隐藏）自己的人，一般都活不久吗？

编辑于 2017-02-02

典型误解

▲ 赞同 1581

▼

收起评论

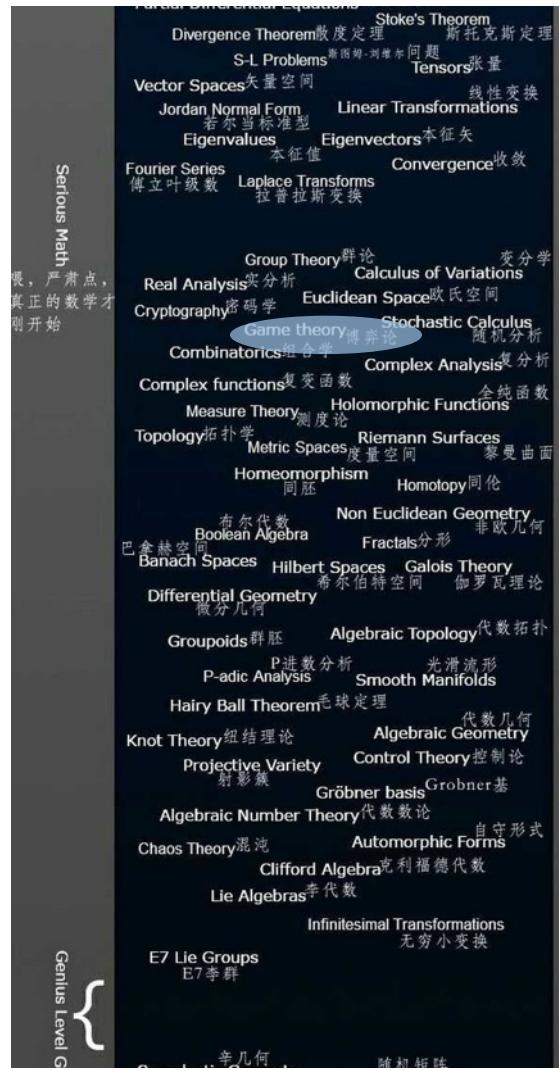
分享

收藏

喜欢

# 什么是博弈论

博弈论本质上是一个数学工具，有一套严密的定义与分析方法



# 什么是博弈论

囚徒困境是最为人熟知的博弈论的例子，

学霸关于每周末是否学习的博弈

——选择学习带来的时间损耗设为40

——期末拿唯一奖学金带来的收益设为100



学霸甲



学霸乙

		不学习	学习
不学习	50, 50	0, 60	
	60, 0	10, 10	

# 什么是博弈论

如果你是学霸甲，如何做决策？

- 若学霸乙选择“不学习”
- 若学霸乙选择“学习”



学霸甲



学霸乙

		学霸乙	
		不学习	学习
不学习		50, 50	0, 60
学习		60, 0	10, 10

# 什么是博弈论

学霸甲和学霸乙不约而同都选择“学习”

——这是双输的结果，因为如果约定好都选

择“不学习”，双方收益都将增加

——粗略解释内卷成因



学霸甲



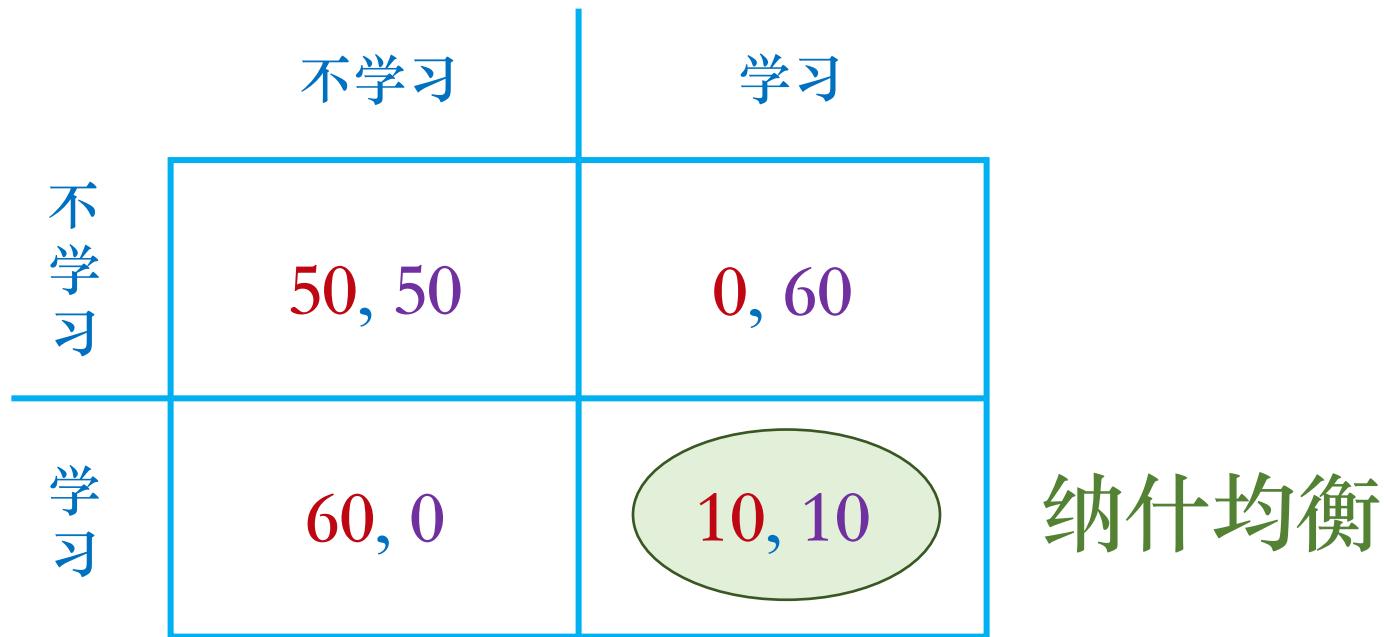
学霸乙

		不学习	学习
不学习	50, 50	0, 60	
	60, 0	10, 10	

# 什么是博弈论

博弈场景即个体收益受他人决策影响的场景

博弈论研究博弈场景下每个的个体决策：如何预测对方决策、如何优化自身决策



双赢决策（即甲乙都选择“不学习”）不是纳什均衡

# 罚点球的例子



- 射手和守门员几乎同时决定往哪边射门/往哪边扑救
- 如何估算上万次罚点球中射手或守门员的决策?
- 是不是往左射和往右射的比例为1:1?
- 为实现估算你需要知道哪些数据?

# 罚点球的例子

Journal of Artificial Intelligence Research 71 (2021) 41-88

Submitted 11/2020; published 05/2021

## Game Plan: What AI can do for Football, and What Football can do for AI

Karl Tuyls, Shayegan Omidshafiei, Paul Muller, Zhe Wang, Jerome Connor,  
Daniel Hennes

KARLTUYLS@ SOMIDSHAFIEI@ PMULLER@ ZHEWANG@ JEROMECONNOR@ HENNES@ ...GOOGLE.COM  
*DeepMind*

Ian Graham, William Spearman, Tim Waskett, Dafydd Steele  
IAN.GRAHAM@ WILLIAM.SPEARMAN@ TIM.WASKETT@ DAFYDD.STEELE@ ...LIVERPOOLFC.COM  
*Liverpool Football Club*

Pauline Luc, Adria Recasens, Alexandre Galashov, Gregory Thornton, Ro-  
muald Elie, Pablo Sprechmann, Pol Moreno, Kris Cao, Marta Garnelo, Pra-  
neet Dutta, Michal Valko, Nicolas Heess, Alex Bridgland, Julien Pérolat, Bart  
De Vylder, S. M. Ali Eslami, Mark Rowland, Andrew Jaegle, Remi Munos,  
Trevor Back, Razia Ahamed, Simon Bouton, Nathalie Beauguerlange, Jack-  
son Broshear, Thore Graepel, Demis Hassabis

PAULINELUC@ ARECASENS@ AGALASHOV@ GREGTHORNTON@ RELIE@ PSPRECHMANN@ POLC@  
KRISCAO@ GARNELO@ PRANEETDUTTA@ VALKOM@ HEESS@ ABRIDGLAND@ PEROLAT@ BARTDV@  
AESLAMI@ MARKROWLAND@ DREWJAEGLE@ MUNOS@ BACK@ RAHAMED@ BOUTON@ NBEAU-  
GUERLANGE@ BROSHEAR@ THORE@ DEMISHASSABIS@ ...GOOGLE.COM  
*DeepMind*

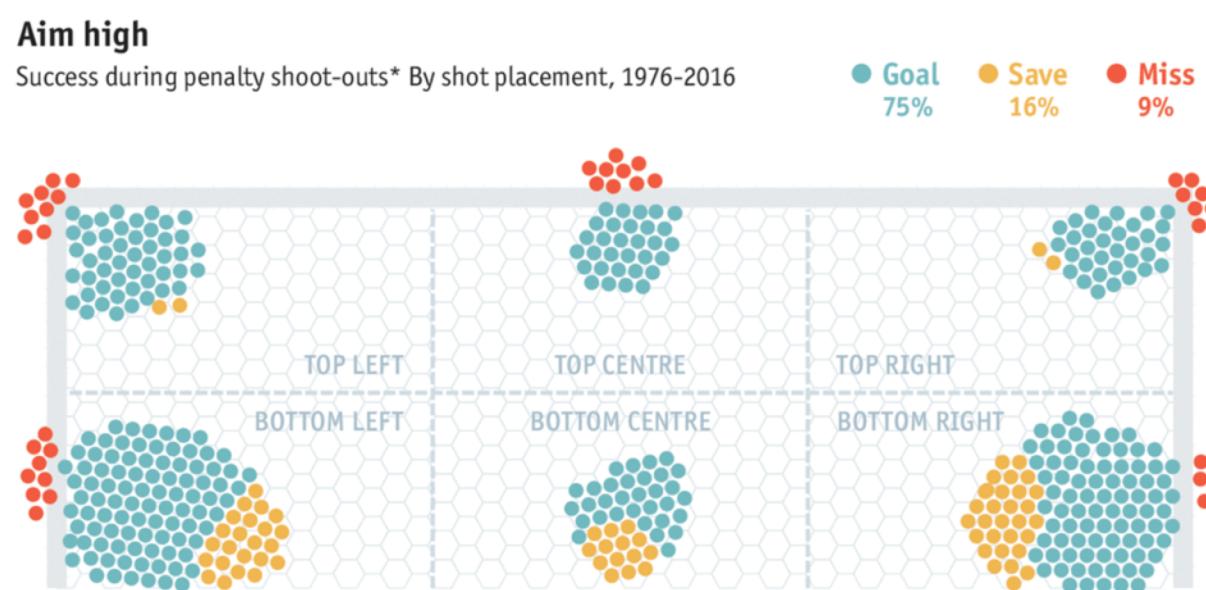
以下例子引自DeepMind和Liverpool足球俱乐部最近发表的文章

# 罚点球的例子

利用博弈论估算射手/守门员决策的分布

——需要对博弈 (Game) 进行建模 (定义决策、定义收益)

► 将射手/守门员的可用决策定义为集合{左, 中, 右}?



# 罚点球的例子

利用博弈论估算射手/守门员决策的分布

——需要对博弈 (Game) 进行建模 (定义决策、定义收益)

- 将射手/守门员的可用决策定义为集合{惯用脚方向、非惯用脚方向}?  
(为简化, 将中间方向归为惯用脚方向)

		守门员	
		射手惯用方向	非射手惯用向
射手	射手惯用方向	? , ?	? , ?
	非射手惯用向	? , ?	? , ?

# 罚点球的例子

利用博弈论估算射手/守门员决策的分布

——需要对博弈（Game）进行建模（定义决策、定义收益）

► 定义收益：分数的期望值

		守门员	
		射手惯用方向	非射手惯用向
射手	射手惯用方向	? , ?	? , ?
	非射手惯用向	? , ?	? , ?

# 罚点球的例子

利用博弈论估算射手/守门员决策的分布

——需要对博弈（Game）进行建模（定义决策、定义收益）

► DeepMind的文章分析真实比赛（各大杯赛、联赛）中上万次点球得分情况

	守门员	
射手	射手惯用方向	非射手惯用向
	0.704, -0.704	0.907, -0.907
非射手惯用向	0.894, -0.894	0.640, -0.640

# 罚点球的例子

利用博弈论估算射手/守门员决策的分布:

有百分之多少的射手会往惯用方向射，百分之多少守门员会往这个方向扑？

		守门员	
		射手惯用方向	非射手惯用向
射手	射手惯用方向	0.704, -0.704	0.907, -0.907
	非射手惯用向	0.894, -0.894	0.640, -0.640

# 罚点球的例子

		守门员
	射手惯用方向	非射手惯用向
射手	0.704, -0.704 0.894, -0.894	0.907, -0.907 0.640, -0.640

论文中数据（未检验）

混合策略纳什均衡

射手 56.9% 惯用向； 43.1% 非惯向

守门员 59.2% 惯用向； 40.8% 非惯向

球赛真实数据

射手 52.5% 惯用向； 47.5% 非惯向

守门员 61.5% 惯用向； 38.5% 非惯向

# 罚点球的例子

## 混合策略纳什均衡

射手 56.9% 惯用向； 43.1% 非惯向      守门员 59.2% 惯用向； 40.8% 非惯向

## 球赛真实数据

射手 52.5% 惯用向； 47.5% 非惯向      守门员 61.5% 惯用向； 38.5% 非惯向

- ▶ 考虑到模型比较粗糙（未区分往中路射、未考虑有经验的守门员的预判、未区分不同球员等），数据规模不大。预测结果已然很不错。
- ▶ 绝大多数球员不会用博弈论分析决策，但事实结果和博弈论分析相对吻合，怎么解释？

# 什么是博弈论



研究博弈场景下个体决策（如何预测他人决策、优化自身决策）的数学方法论



研究的**博弈问题**包括：

棋牌：德州扑克

交通：车辆的路径选择

通信：移动用户网络选择

安全：网络攻防博弈

市场：互联网公司博弈

电商：交易机器人  
等等



AlphaGo之后，机器在棋牌游戏上战胜人类的又一重要事件

# 背景介绍

什么是博弈论



什么是机制设计

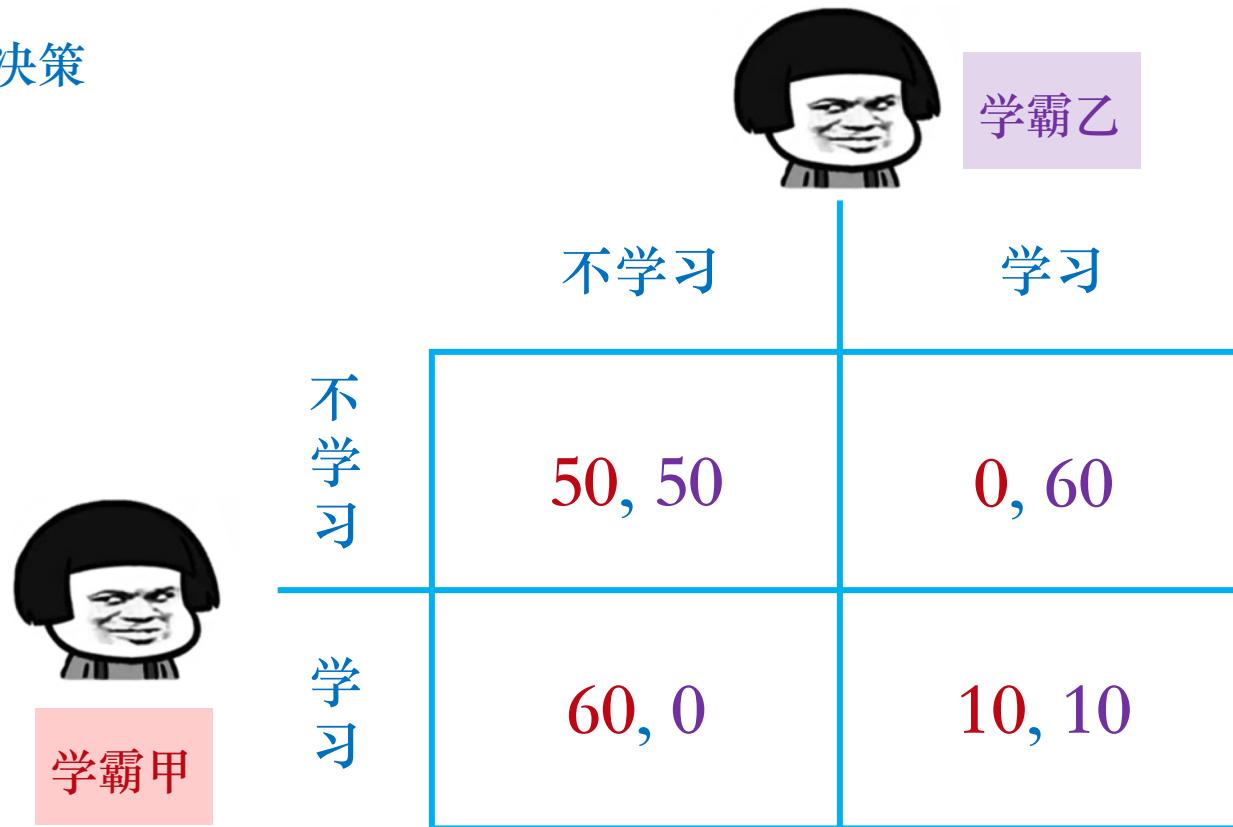
完全理性假设

# 什么是机制设计

机制设计（mechanism design）可以看作博弈论的反问题：

博弈论：在给定博弈规则下，分析参与博弈的个体的决策

机制设计：作为系统管理者，设计博弈规则，引导参与博弈的个体选择系统管理者所期望的决策



# 什么是机制设计

机制设计可以看作博弃论的反问题：

博弃论：在给定博弃规则下，分析参与博弃的个体的决策

机制设计：作为系统管理者，设计博弃规则，引导参与博弃的个体选择系统管理者所期望的决策

原规则：仅第一名拿100

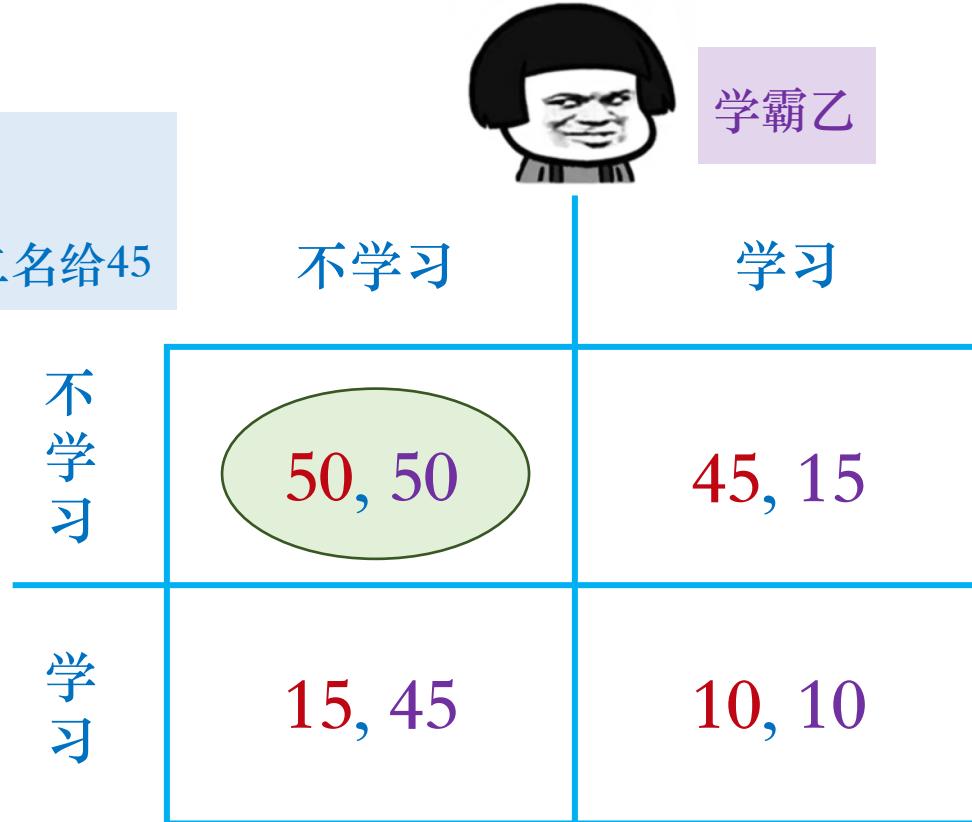
新规则：第一名给55、第二名给45



学霸甲

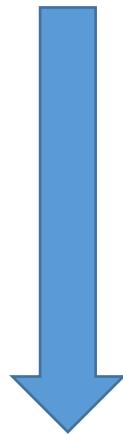


学霸乙



# 什么是机制设计

原规则：仅第一名拿100



新规则：

第一名给55、第二名给45

	不学习	学习
不学习	50, 50	0, 60
学习	60, 0	10, 10

	不学习	学习
不学习	50, 50	45, 15
学习	15, 45	10, 10

# 失败的机制

---

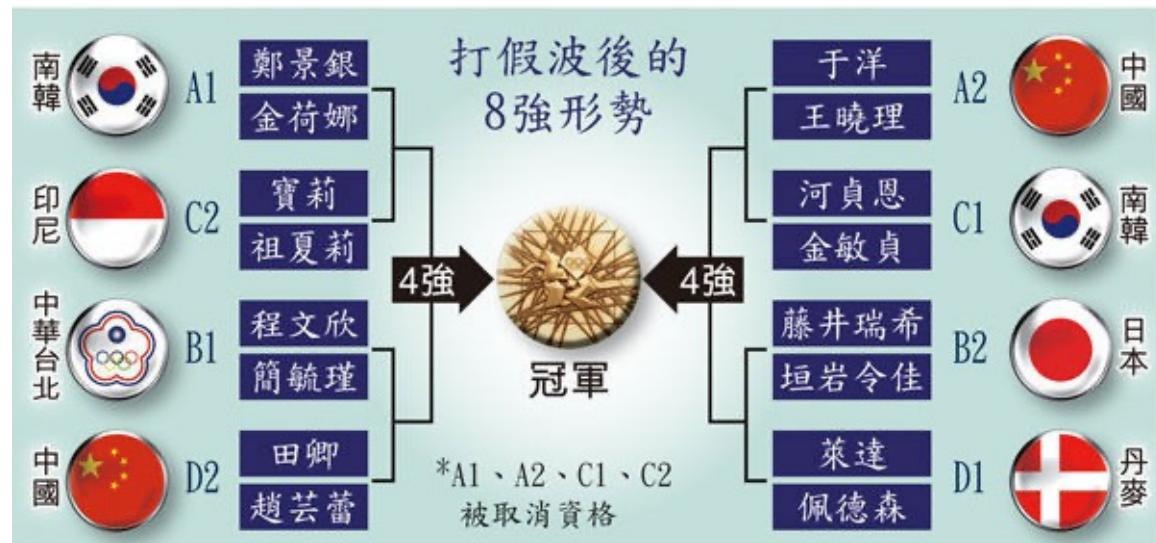
2012年伦敦奥运会羽毛球女双选手被取消参赛资格事件

羽联通告取消于洋王晓理等4对女双参赛资格

<http://2012.sina.com.cn> 2012年08月01日20:56 新浪体育 微博

# 失败的机制

当时另一队中国选手（田/赵）已经以小组第二晋级淘汰赛，于/王还有一场小组赛要打。如果于/王赢得比赛，会以小组第一晋级，在淘汰赛提前遭遇田/赵，这样中国队无法包揽金银牌；如果于/王输掉比赛，会以小组第二晋级，中国队内部不会提前遭遇、有可能包揽金银牌。



# 什么是机制设计

---



研究如何设计机制从而影响个体的收益函数，进而影响博弈场景中个体的决策，最终实现优化系统性能的目的



机制设计场景包括：

通信：无线频谱拍卖

电商：卖家排序

网络：P2P共享方案

交通：共享出行补贴定价

财政：税收方案

金融：保险条目设置

等等

# 背景介绍

什么是博弈论

什么是机制设计

完全理性假设



# 美数猜博奕 (Beauty Contest Game)

---

参与博弈的每个人猜一个0到100之间的整数。

汇总所有人猜的数字，计算这些数字的平均值的 $2/3$ ，猜测数字离这个值最近的玩家获胜。

- ➡ 这个博弈的纳什均衡策略是什么？
- ➡ 纽约时报曾请61140个读者参与这个游戏，如果你是第61141个读者，你会猜什么数字？

# 美数猜博奕 (Beauty Contest Game)

参与博弈的每个人猜一个0到100之间的整数。

汇总所有人猜的数字，计算这些数字的平均值的 $2/3$ ，猜测数字离这个值最近的玩家获胜。

► 这个博弈的纳什均衡策略是什么？

第一个纳什均衡：所有人猜1

第二个纳什均衡：所有人猜0

# 美数猜博奕 (Beauty Contest Game)

---

参与博弈的每个人猜一个0到100之间的整数。

汇总所有人猜的数字，计算这些数字的平均值的 $2/3$ ，猜测数字离这个值最近的玩家获胜。

- ➡ 这个博弈的纳什均衡策略是什么？
- ➡ 纽约时报曾请61140个读者参与这个游戏，如果你是第61141个读者，你会猜什么数字？猜0或者1能获胜吗？

# 美数猜博弈 (Beauty Contest Game)

PERCENT OF READERS PICKING EACH NUMBER:



第一个纳什均衡：所有人猜1；第二个纳什均衡：所有人猜0

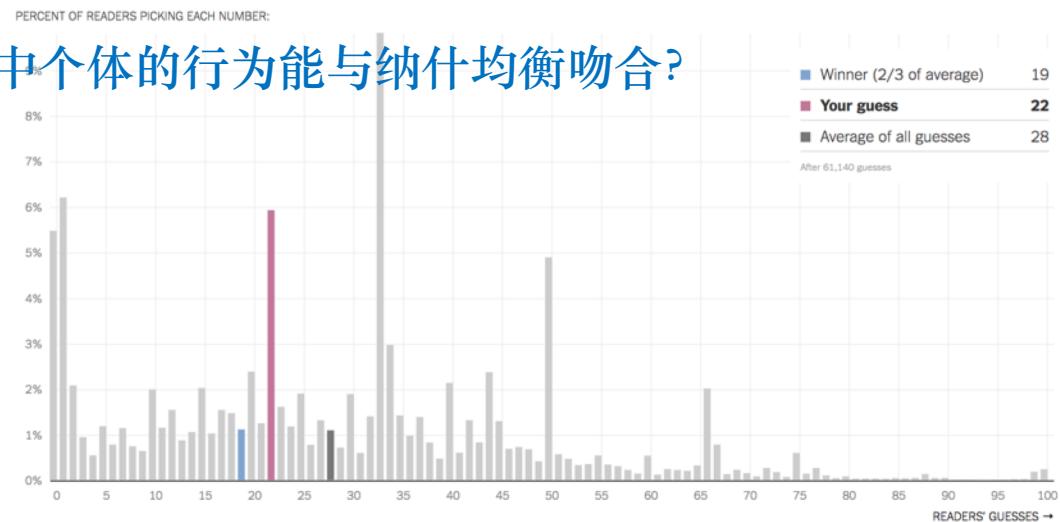
实际结果和纳什均衡不吻合，为什么？什么情况下能够吻合？

# 完全理性假设

纳什均衡（所有人猜1或所有人猜0）与实际61140个玩家的结果不吻合  
——原因是实际中参与博弈的个体**不具备完全理性**

许多研究假设参与博弈或机制设计场景中的个体具备完全理性，这导致研究结果（对个体行为的预测或设计的机制）的实用性受到影响。

什么情况下美数猜博弈中个体的行为能与纳什均衡吻合？



# 主要内容

---



## 背景介绍

- ▶ 什么是博弈论
- ▶ 什么是机制设计
- ▶ 完全理性假设



## 与AI的结合

- ▶ 博弈论 + AI
- ▶ 机制设计 + AI

# 与AI的结合

博弈论 + AI

机制设计 + AI

# 广告位拍卖

---

## ▶ 各类拍卖是机制设计在实际中最广泛的应用

频谱拍卖：如在2016-2017的FCC激励拍卖中，美国政府净盈利近百亿美元

广告位拍卖：Google在2020年的广告盈利为上千亿美元

## ▶ 围绕拍卖理论授予的诺贝尔经济学奖

1996: William Vickrey (**Vickrey auction**) , James Mirrlees

2002: Vernon L. Smith (**combinatorial auction**) , Daniel Kahneman

2007: Roger Myerson (**optimal auction**) , Leonid Hurwicz, Eric S. Maskin

2020: Paul Milgrom (**auction theory**) , Robert B. Wilson (**auction theory**)

# 广告位拍卖

► 2021年百度第一季度广告收入163亿元

Baidu 百度 手机 百度一下

网页 图片 贴吧 新闻 视频 知道 音乐 地图 文库 更多»

百度为您找到相关结果约100,000,000个

▼搜索工具

**手机【天猫手机馆】高清大屏\_省钱更省心** 广告

手机-天猫手机馆,品牌旗舰,强劲配置,高性能低价格,分期免息,超值实惠,正品包邮!网购手机,来...  
品牌: 三星 | OPPO | vivo | 荣耀 | 360 | 魅族 | 中兴 | 天翼 | TOPEAK | 更多»



[查看更多相关信息>>](#)  
[shouji.tmall.com](http://shouji.tmall.com) 2018-01 ▼  - 6790条评论

**手机比较\_苏宁手机年货节\_爆款惠不停\_新手机陪你过新年**

  
苏宁易购 suning.com

苏宁易购手机比较,正品低价,购苏宁超市产品领大额手机券,480减50,99减200,惊喜不断!苏宁易购手机比较,支持货到付款,...  
机身内存: 64GB 128GB 32GB 8GB 256GB 16GB  
价格: 0-1000 1000-2600 2600-3499 3499-4600 4600以上  
[cuxiao.suning.com](http://cuxiao.suning.com) 2018-01 ▼  - 2164条评论

**JD 手机 购机到【京东】JD.COM**

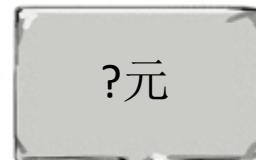
  
手机品类齐全,到京东轻松购物,购手机手快有 手慢也有!京东正品行货,服务100分,送货到家,靠谱!  
[www.jd.com](http://www.jd.com) 2018-01 ▼  - 5068条评论

4个广告位

# 广告位拍卖



百度



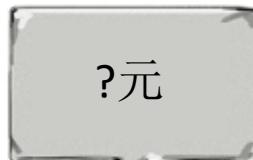
?元

华为



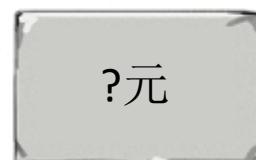
?元

苹果



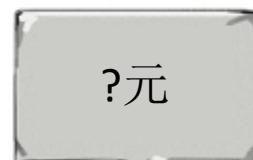
?元

小米



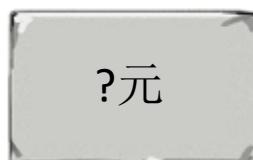
?元

OPPO



?元

三星

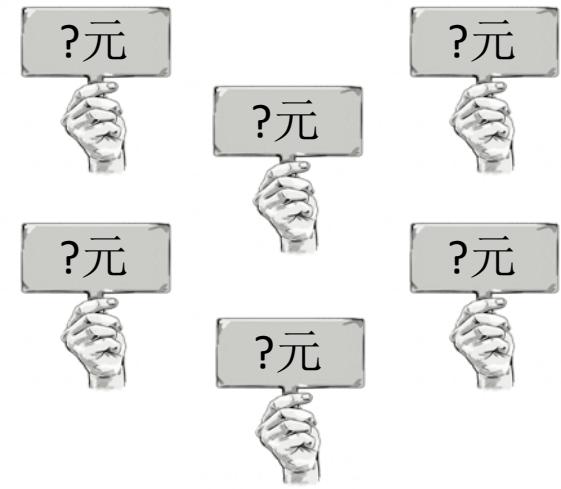


?元

vivo

# 广告位拍卖

- ▶ 百度设计拍卖机制  
⇒ 广告商决定竞拍价（博弈）
- ▶ 不同拍卖机制导致不同竞拍行为
- ▶ 百度希望广告商竞价越高越好



拍卖机制设计：

- 分配规则：如何把广告位分给广告商？（不一定竞价高就一定分配最好的广告位）
- 支付规则：如何向赢得广告位的广告商收费？（不一定按照竞价进行收费）
- \* 是否以及如何设置拍卖底价

# 广告位拍卖

---

- ▶ 设置**拍卖底价**可以显著刺激广告商竞价行为，提升平台收益

*Proceedings > EC '11 > Reserve prices in internet advertising auctions: a field experiment*

ABSTRACT

## **Reserve prices in internet advertising auctions: a field experiment**



**Authors:**  Michael Ostrovsky,  Michael Schwarz [Authors Info & Affiliations](#)

Yahoo!在与461648个关键词相关的广告位进行拍卖实验，结果说明  
通过添加（未经优化的）**拍卖底价**可以将利润提升3.8%

# 如何设置拍卖底价

对单件物品的拍卖，Roger Myerson（诺奖得主）根据博弈论计算的最优拍卖机制

## Definition (virtual valuation)

Bidder  $i$ 's virtual valuation is  $\psi_i(v_i) = v_i - \frac{1-F_i(v_i)}{f_i(v_i)}$ .

Let us assume this is increasing in  $v_i$  (e.g., for a uniform distribution it is  $2v_i - 1$ ).

## Definition (bidder-specific reserve price)

Bidder  $i$ 's bidder-specific reserve price  $r_i^*$  is the value for which  $\psi_i(r_i^*) = 0$ .

设置拍卖底价的方法

## Theorem (Myerson (1981))

The optimal (single-good) auction in terms of a direct mechanism: The good is sold to the agent  $i = \arg \max_i \psi_i(\hat{v}_i)$ , as long as  $v_i \geq r_i^*$ . If the good is sold, the winning agent  $i$  is charged the smallest valuation that he could have declared while still remaining the winner:

$$\inf\{v_i^* : \psi_i(v_i^*) \geq 0 \text{ and } \forall j \neq i, \psi_i(v_i^*) \geq \psi_j(\hat{v}_j)\}.$$

- 理论假设所有竞价者具备完全理性
- 广告位拍卖是更复杂的多物品拍卖

# 如何设置拍卖底价

## Reinforcement Mechanism Design: With Applications to Dynamic Pricing in Sponsored Search Auctions\*

Weiran Shen<sup>1</sup>, Binghui Peng<sup>1</sup>, Hanpeng Liu<sup>1</sup>, Michael Zhang<sup>2</sup>,  
Ruohan Qian<sup>3</sup>, Yan Hong<sup>3</sup>, Zhi Guo<sup>3</sup>, Zongyao Ding<sup>3</sup>, Pengjun Lu<sup>3</sup>, Pingzhong Tang<sup>1†</sup>

<sup>1</sup>Institute for Interdisciplinary Information Sciences, Tsinghua University

<sup>2</sup>Chinese University of Hong Kong

<sup>3</sup>Baidu Inc.

AAAI'20

## 利用AI方法以及历史数据设置拍卖底价

- 适用于有限理性竞拍者
- 适用于多物品拍卖
- 百度线上实验

# 如何设置拍卖底价

步骤

预测众多广告商竞拍行为



根据预测设置拍卖底价

本质

有监督学习



动态决策

方法

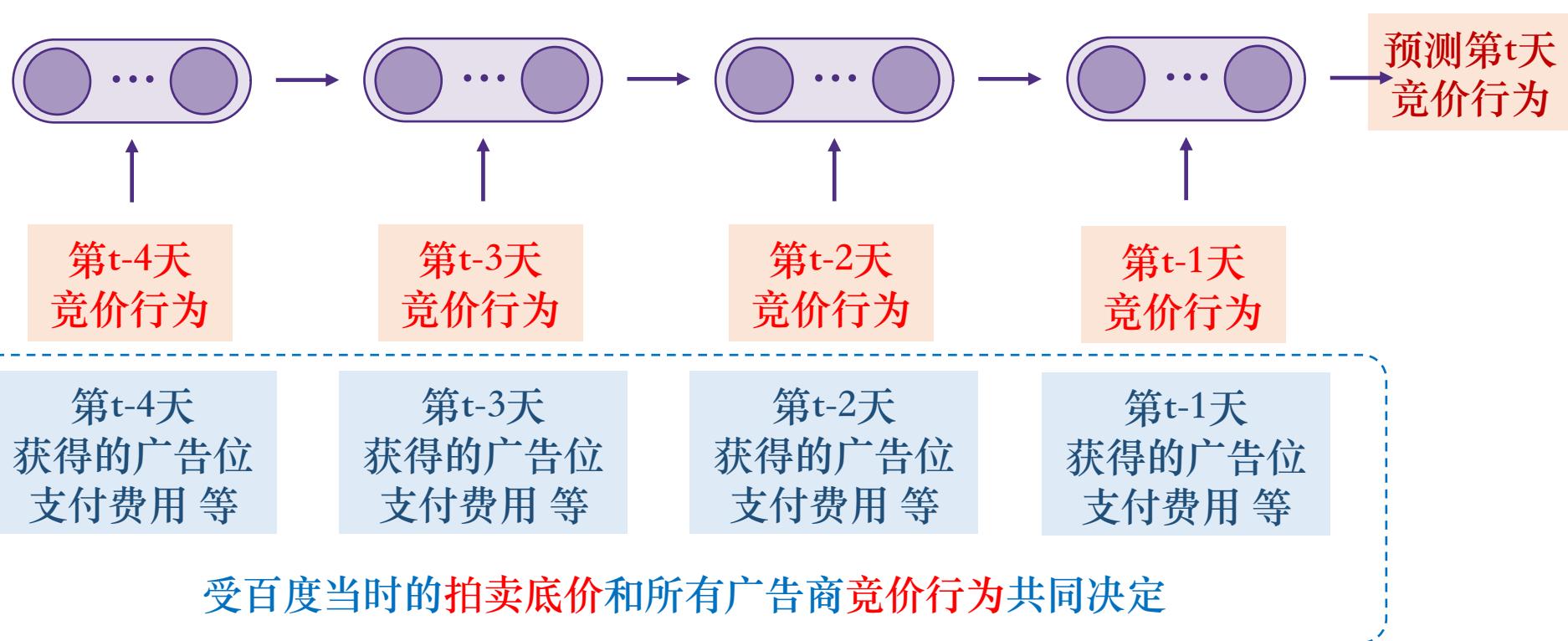
循环神经网络



蒙特卡洛树搜索

# 预测竞拍行为

为每个广告商建立一个LSTM模型，根据前4天信息预测接下来1天行为



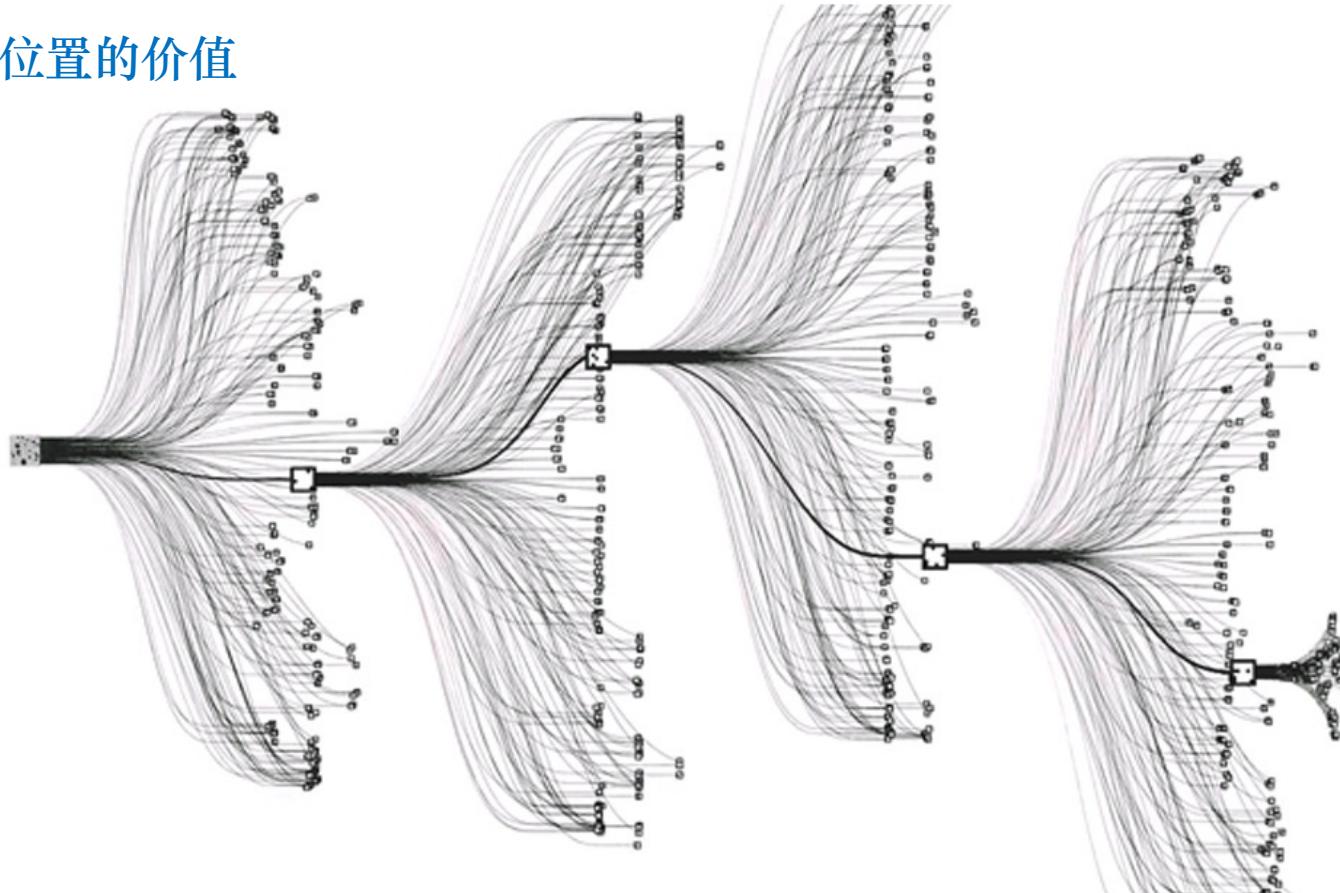
# 根据预测设置拍卖底价

---

- ➡ 当前时刻t的**拍卖底价**会影响：
  - 当前时刻t广告收益
  - $t+1$ 时刻广告商的竞价,  $t+1$ 时刻广告收益
  - $t+2$ 时刻广告商的竞价,  $t+2$ 时刻广告收益
  - ...
- ➡ 用**蒙特卡洛树搜索**动态决定当前时刻的**拍卖底价**
  - 被用于设计AlphaGo, 解决围棋中的动态决策问题

# 根据预测设置拍卖底价

AlphaGo对每一个可能落子的位置通过蒙特卡洛树搜索预演后续变化来估测该落子位置的价值



类似，可以对每一种拍卖底价的取值通过蒙特卡洛树搜索和对广告商行为预测模型预演后续变化，估测该拍卖底价的价值。

# 内容小结

---



## 背景介绍

- ▶ 什么是博弈论
- ▶ 什么是机制设计
- ▶ 完全理性假设



## 与AI的结合

- ▶ 博弈论 + AI
- ▶ 机制设计 + AI

# 谢谢！

