

# 第三部分：简单决策系统

章宗长

2021年4月14日

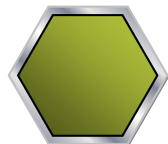
# 内容安排



效用理论



决策网络



信息价值



专家系统



单步博弈

# 信息价值

- 至此，在决策网络中，我们假设仅观察到 $o_1^1$ ，基于这一正面的诊断测试结果，决定是否治疗
- 通过执行其他的诊断测试来降低因误诊而延误治疗的风险
- 可以通过计算**信息价值**来决定执行哪种诊断测试
  - 一条给定信息的价值：获取该信息之后和之前的最优行动的期望价值之间的差
- 信息价值理论
  - 使得Agent能够选择要获取什么信息
  - 涉及序贯决策的一种简化形式，即观察行动只影响Agent的信念状态，而不是外在的物理状态



# 一个简单实例

## ■ 假设：

- 一家石油公司想购买不可区分的 $n$ 块海洋开采权中的一块
- 仅有一块含有价值 $C$ 美元的石油，其他块是没有价值的
- 每块的标价是 $C/n$ 美元
- 该公司是风险中立的
- 一个地震学家为该公司提供对第3块的调查结果，结果明确指出这块海洋是否含有石油

## ■ 问：该公司应该愿意为这条信息支付多少钱？



## 一个简单实例（续）

- 考察如果公司得到这条信息将会做什么：
  - 调查结果以 $1/n$ 的概率指出第3块海洋中含有石油

将以 $\frac{C}{n}$ 美元买下第3块海洋开采权，获利：

$$C - \frac{C}{n} = \frac{(n-1)C}{n}$$

- 调查结果以 $(n-1)/n$ 的概率指出第3块海洋中不含有石油

将以 $\frac{C}{n}$ 美元买下不同的另一块，期望获利：

$$\frac{C}{n-1} - \frac{C}{n} = \frac{C}{n(n-1)}$$

- 给定调查信息，可以计算期望利润：

$$\frac{1}{n} \times \frac{(n-1)C}{n} + \frac{n-1}{n} \times \frac{C}{n(n-1)} = \frac{C}{n}$$

公司愿意为这条信息最多支付 $C/n$ 美元

# 信息价值的通用公式

- $EU^*(\mathbf{o})$ : 给定一组观察 $\mathbf{o}$ , 最优行动对应的期望效用
- 给定一组观察 $\mathbf{o}$ , 变量 $O'$ 的期望价值:

$$VOI(O' | \mathbf{o}) = \underbrace{\left( \sum_{o'} P(o' | \mathbf{o}) EU^*(\mathbf{o}, o') \right)}_{\text{获取变量 } O' \text{ 的观察之后的最优行动的期望效用}} - \underbrace{EU^*(\mathbf{o})}_{\text{获取变量 } O' \text{ 的观察之前的最优行动的期望效用}}$$

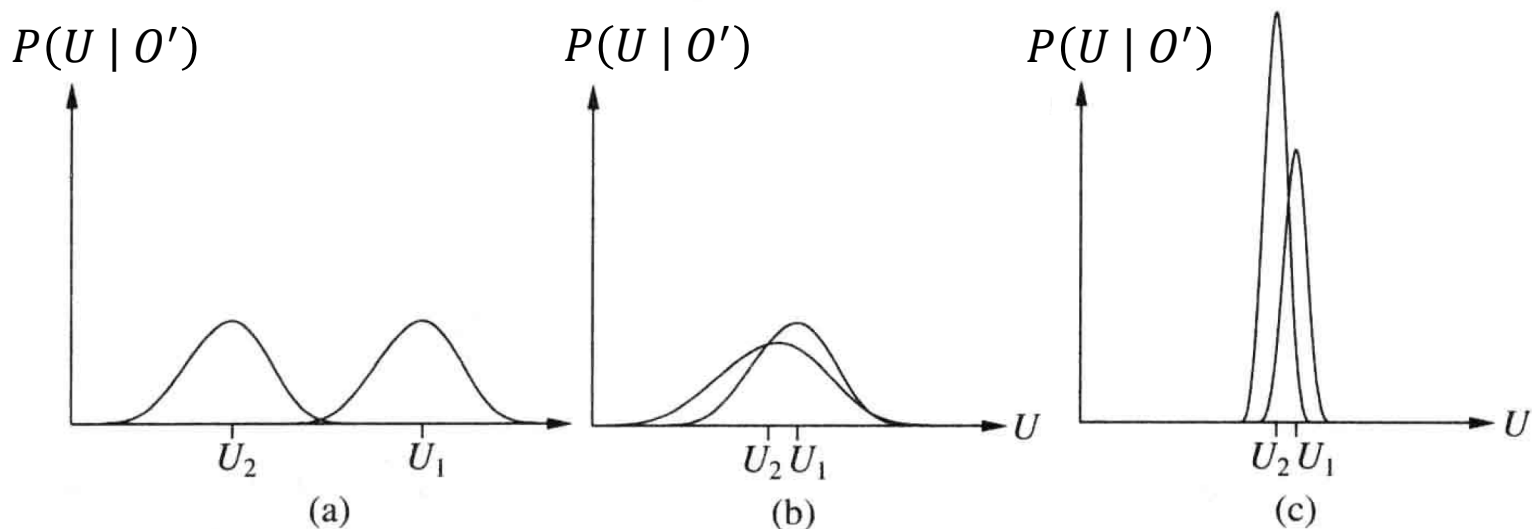
获取变量 $O'$ 的观察之后的  
最优行动的期望效用

获取变量 $O'$ 的观察之前的  
最优行动的期望效用

- 只有在变量 $O'$ 的观察导致不同的最优决策时,  $VOI(O' | \mathbf{o})$ 才大于0
- 若不论诊断测试为何种结果, 最优决策均为治疗疾病, 则观察测试结果的价值为0

# 信息价值的三种一般情况

- 考虑只有两个行动 $a_1$ 和 $a_2$ ；这两个行动的当前期望效用是 $U_1$ 和 $U_2$
- 信息 $O' = o'$ 将为行动产生某些新的期望效用 $U'_1$ 和 $U'_2$



- 信息价值的三种一般情况：
  - 在 (a) 中， $a_1$  几乎肯定地一直好于  $a_2$ ，因此不需要信息
  - 在 (b) 中，选择并不清楚，信息至关重要
  - 在 (c) 中，选择也不清楚，但因为选择没有多少区别，所以信息的价值较小

# 信息价值的属性

- 定理：信息的期望价值是**非负的**：

$$VOI(O_j | \mathbf{o}) \geq 0 \quad \forall \mathbf{o}, O_j$$

关于期望信息而不是真实价值的定理

- 如果信息碰巧具有误导性，那么额外的信息容易导致计划比原有的计划更差

- 信息价值是**不可累加的**：

$$VOI(O_i, O_j | \mathbf{o}) \neq VOI(O_i | \mathbf{o}) + VOI(O_j | \mathbf{o}) \quad \text{一般情况下}$$

- 信息价值是**独立于次序的**：

$$\begin{aligned} VOI(O_i, O_j | \mathbf{o}) &= VOI(O_i | \mathbf{o}) + VOI(O_j | O_i, \mathbf{o}) \\ &= VOI(O_j | \mathbf{o}) + VOI(O_i | O_j, \mathbf{o}) \end{aligned}$$



# 信息收集的成本

- 信息价值仅捕获在观察一个变量后期望效用的增益
- 实际上，需要花费**成本**来获得一个变量的观察
  - 一些诊断测试是便宜的
    - 如：测体温
  - 另一些诊断测试是成本高昂的
    - 如：腰椎穿刺
  - 在选择测试方案时，既要考虑信息价值，也要考虑测试成本



# 信息收集Agent的实现

- (1) 把观察到的机会结点实例化为证据变量
  - (2) 计算未观察变量的信息价值和成本
  - (3) 选信息价值与成本之差最大的未观察变量为下一个观察
  - (4) while 所选观察的信息价值与成本之差大于0
    - (a) 执行步骤 (1 ~ 3)
  - (5) 用评价决策网络的算法选择最优行动
- 这里使用贪心法得到观察序列
    - 启发式方法
    - 近视的 (myopic) : 所获得的序列不一定是最优观察序列

# 小结：信息价值

## ■ 信息价值

- 获取信息之后和之前的最优行动的期望价值之间的差
- 有时需要花费成本来获得一个变量的观察

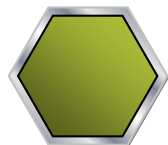
## ■ 信息价值的属性

- 非负、不可累加、独立于次序

## ■ 信息收集Agent的实现

- 贪心法

# 内容安排



效用理论



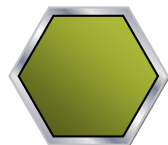
决策网络



信息价值



专家系统



单步博弈

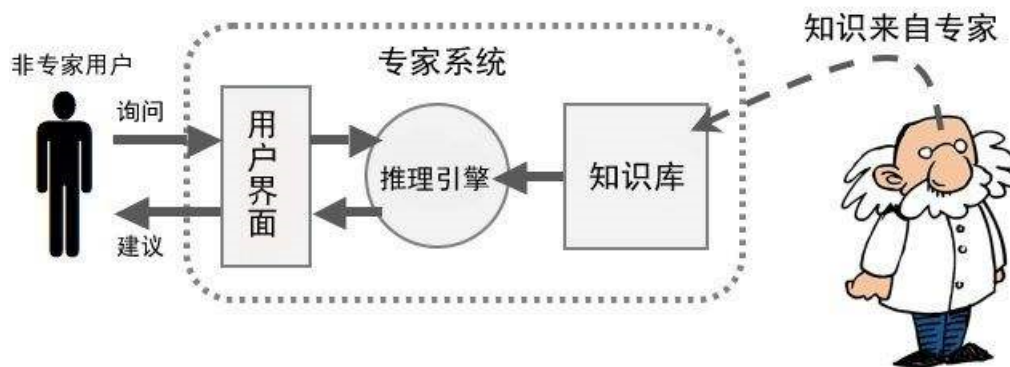
# 专家系统

- 专家系统概述
- 专家系统的结构
- 决策专家系统

# 专家系统

- 一类具有专门知识和经验的计算机智能程序系统
- 模拟人类专家解决领域问题的能力
- 基于知识的系统：知识就是力量

**专家系统=知识库+推理机**

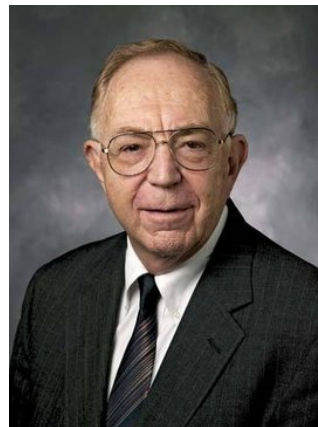


# 专家系统（续）

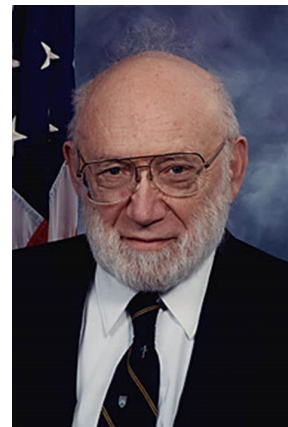
- 一个专家系统应该具备以下3个要素：
  - 具备某个应用领域的专家级知识
  - 能模拟专家的思维
  - 能达到专家级的解题水平
- 知识工程：建造一个专家系统的过程
  - 把软件工程的思想应用于设计基于知识的系统
- 知识工程包括以下4个方面：
  - 知识获取
  - 知识表示
  - 软件设计
  - 编程实现

# 专家系统的发展史

- 1965年，斯坦福大学的计算机科学家费根鲍姆和化学家勒德贝格合作研制了DENDRAL系统
  - 推断化学分子结构
- 20世纪70年代，专家系统的观点逐渐被人们接受，许多专家系统相继研发成功
  - 代表性的有：医药专家系统MYCIN、探矿专家系统PROSPECTOR等
- 20世纪80年代，专家系统的开发趋于商业化，创造了巨大的经济效益



费根鲍姆



勒德贝格



# 专家系统的发展史（续）

- 1977年，中国科学院自动化研究所研制成功了我国第一个专家系统——“中医肝病诊治专家系统”
- 1985年，中国科学院合肥智能机械研究所研制成功了我国第一个农业专家系统——“砂姜黑土小麦施肥专家咨询系统”
- 中国科学院计算技术研究所和中国水产科学研究院东海水产研究所等合作，研制了东海渔场预报专家系统
- 20世纪80年代以来，涌现出了不少专家系统开发工具
  - 国外：EMYCIN、CLIPS等
  - 国内：天马、雄风、OKPS等

# 专家系统的特点

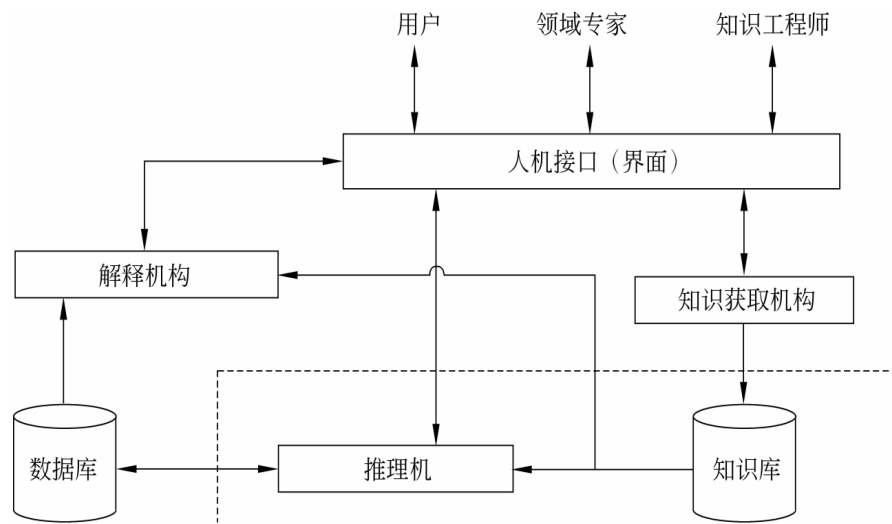
- **启发性**：专家系统能运用专家的知识与经验进行推理、判断和决策
- **透明性**：专家系统能够解释本身的推理过程和回答用户提出的问题，以便让用户能够了解推理过程，提高对专家系统的信赖感
- **灵活性**：专家系统能不断地增长知识，修改原有知识，不断更新

# 专家系统

- 专家系统概述
- 专家系统的结构
- 决策专家系统

# 专家系统的结构

- 专家系统通常由**知识库**、**推理机**、**数据库**、**知识获取机构**、**解释机构**和**人机接口**这6个部分构成



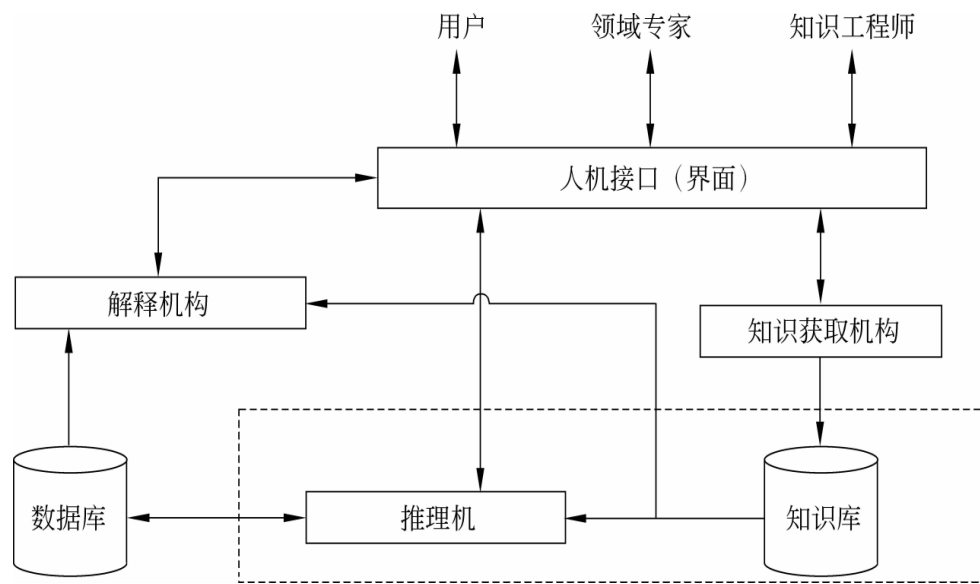
- 知识库**：问题求解所需要的领域知识的集合
- 推理机**：问题求解的核心执行机构

## 基本工作过程

- 用户通过**人机界面**回答系统的提问
- 推理机**将用户输入的信息与**知识库**中的知识进行推理
- 推理得到的中间结果放到**数据库**
- 通过**解释机构**把得出的最终结论呈现给用户

# 专家系统的结构（续）

- **知识获取机构**：负责建立、修改和扩充知识库
  - 知识获取的途径：手工、自动、半自动
- **人机接口**：系统与用户进行交流时使用的界面
- **数据库**：反映当前问题求解状态的集合
  - 既是推理机选用知识的依据，也是解释机构获取推理路径的来源
- **解释机构**：对求解过程作出说明，并回答用户的提问
  - 让用户理解程序在做什么和为什么这样做



# 应用案例：动物识别专家系统

- 动物识别专家系统：识别金钱豹、虎、长颈鹿、斑马、企鹅、鸵鸟、信天翁等7种动物
- 知识库：15条规则（规则1~8）
  - 规则1: IF 动物有毛发 THEN 动物是哺乳动物
  - 规则2: IF 动物有奶 THEN 动物是哺乳动物
  - 规则3: IF 动物有羽毛 THEN 动物是鸟
  - 规则4: IF 动物会飞 AND 会下蛋 THEN 动物是鸟
  - 规则5: IF 动物吃肉 THEN 动物是食肉动物
  - 规则6: IF 动物有犬齿 AND 有爪 AND 眼盯前方 THEN 动物是食肉动物
  - 规则7: IF 动物是哺乳动物 AND 有蹄 THEN 动物是有蹄类动物
  - 规则8: IF 动物是哺乳动物 AND 反刍 THEN 动物是有蹄类动物

# 应用案例：动物识别专家系统（续）

## ■ 知识库：15条规则（规则9~15）

- 规则9: IF 动物是哺乳动物 AND 是食肉动物 AND是黄褐色的 AND 有暗斑点 THEN 动物是金钱豹
- 规则10: IF 动物是黄褐色的 AND 是哺乳动物 AND 是食肉 AND 有黑条纹 THEN 动物是虎
- 规则11: IF 动物有暗斑点 AND 有长腿 AND 有长脖子 AND 是有蹄类动物 THEN 动物是长颈鹿
- 规则12: IF 动物有黑条纹 AND 是有蹄类动物 THEN 动物是斑马
- 规则13: IF 动物有长腿 AND 有长脖子 AND 是黑色的 AND 是鸟 AND 不会飞 THEN 动物是鸵鸟
- 规则14: IF 动物是鸟 AND 不会飞 AND 会游泳 AND 是黑色的 THEN 动物是企鹅
- 规则15: IF 动物是鸟 AND 善飞 THEN 动物是信天翁

## 应用案例：动物识别专家系统（续）

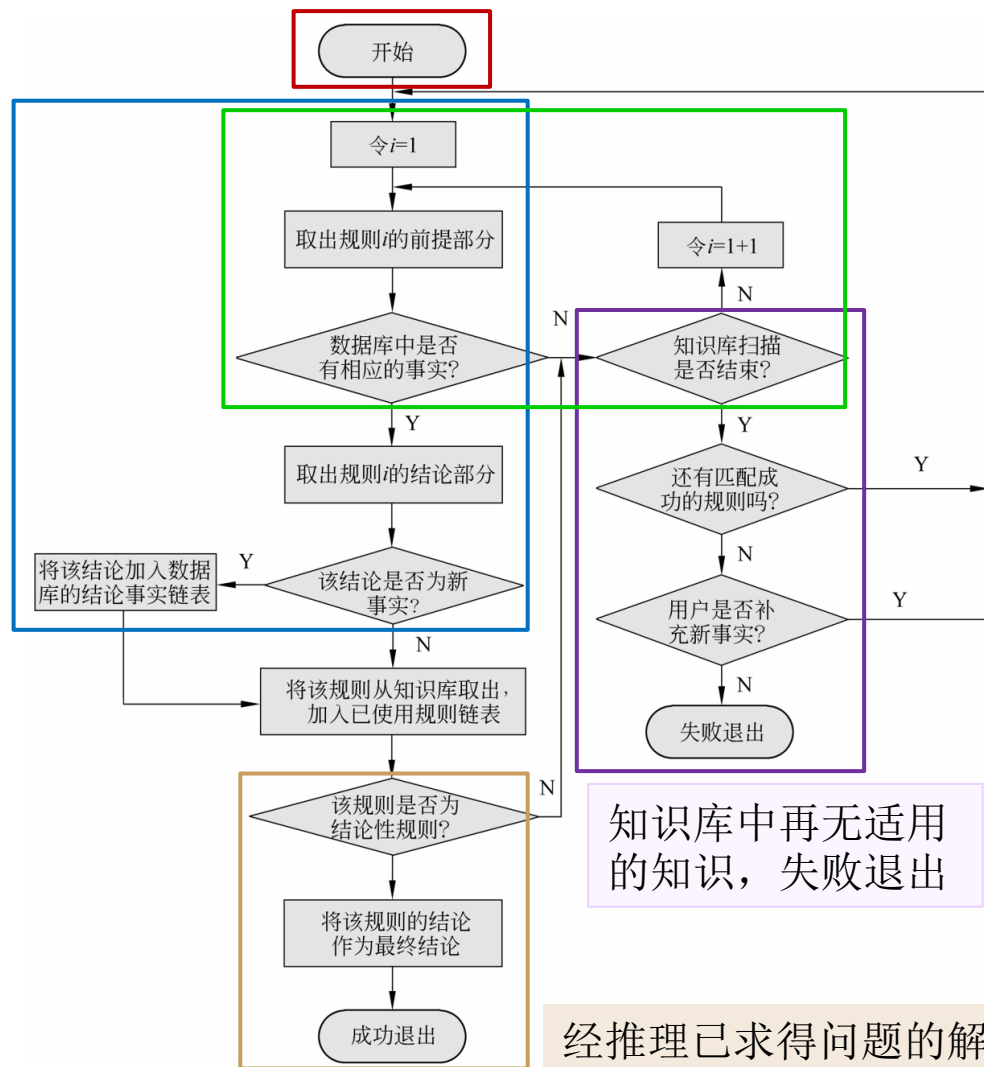
- 在15条规则中，共出现30个概念，也称为**事实**：有毛发、能产奶、哺乳动物、有羽毛、会飞、产蛋、鸟、吃肉、有犬齿、有爪、眼盯前方、食肉动物、有蹄、反刍、有蹄类动物、黄褐色、身上有暗斑点、黑色条纹、有长脖子、有长腿、不会飞、黑白色、会游泳、信天翁、企鹅、鸵鸟、斑马、长颈鹿、老虎、猎豹
- **数据库**：**事实库**，主要存放问题求解的相关信息，包括原始事实、中间结果和最终结论，中间结果又可以作为下一步推理的事实



# 应用案例：动物识别专家系统（续）

■ **推理机**：能执行右图所示推理过程的程序

- 用户首先**初始化**数据库，即把已知事实存放到数据库
- 推理机检查规则库中是否有规则的前提条件可与数据库中已知事实相**匹配**，若有，则把匹配成功的规则的结论部分作为**新的事实**放入数据库
- 检查数据库中是否包含待解决问题的解
  - 是，问题**求解成功**
  - 否，用更新后的数据库中的所有事实**重新进行匹配**



# 应用案例：动物识别专家系统（续）

- **解释机构**：回答系统如何推出最终结论，其功能的实现与推理机密切相关
  - 对推理进行实时跟踪
  - 在推理过程中，每匹配成功一条规则，就记下该规则的序号
  - 推理结束后，把问题求解所使用的规则按次序记录下来，得到整个推理路径

# 专家系统

- 专家系统概述
- 专家系统的结构
- 决策专家系统

# 早期专家系统的缺陷

- 通常使用**条件-行动规则**来推荐行动，而不是使用关于结果偏好的明确表示
- 按照**似然性**的顺序排列可能的诊断，并报告那个最可能的诊断

诊断1：什么病都没有 60%

诊断2：患有重感冒 35%

诊断3：患有肺癌 5%

这可能是灾难性的！

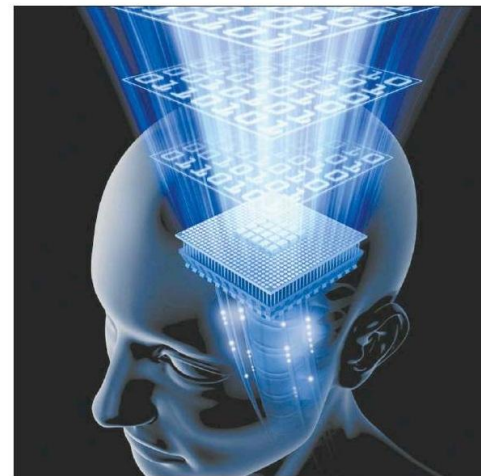


报告最可能的诊断

诊断1：什么病都没有

# 决策专家系统

- 测试或者治疗方案应该同时取决于概率和效用
- 目前的医疗专家系统能够考虑信息价值以推荐测试，然后描绘出不同的诊断
- 决策专家系统：决策网络
  - 贝叶斯网络
    - 从证据产生可靠的概率推理
  - 效用、行动
    - 反映Agent的偏好
    - 推荐最优决策
  - 避免缺陷：把似然性和重要性相混淆



# 创建决策专家系统

## (1) 识别可能行动的空间

- ❑ 避碰系统：爬升、下降、什么都不做
- ❑ 医疗系统：外科手术、心脏支架、药物
- ❑ 在有些问题中，把行动空间因子化为多个决策变量

## (2) 识别与问题相关的观察到的和未观察到的变量

- ❑ 避碰系统：与另一架飞机的相对角度、另一架飞机的真实位置
- ❑ 医疗系统：各种症状、失调

## (3) 识别不同机会结点和决策结点之间的关系

- ❑ 专家判断、结构学习

# 创建决策专家系统（续）

## （4）选择用于表示条件概率分布的模型

- 离散结点：表格表示
- 连续结点：参数模型（如线性高斯），通过专家指定参数或者用参数学习方法从数据中学得

## （5）加入效用结点，添加功能边，用以连接与效用结点相关的机会结点、决策结点

- 效用结点的参数
  - 可由人类专家用**偏好启发式**方法给定
  - 也可**调参**得到，使得由决策网络推出的最优决策与专家决策一致

# 创建决策专家系统（续）

## （6）所得到的决策网络应由人类专家验证和精化

- 如果决策网络与人类专家的决策不一致
  - 检查决策网络，分析为什么不同
  - 有时会要修正条件概率，变量间的关系，或者添加新变量到模型中
  - 有时会修正人类专家的行动选择
- 执行**敏感性分析**，检查最优决策是否对分配的概率和效用的微小变化敏感
  - 若是，则花费更多的资源以收集更好的数据可能是值得的
- 在找到一个合适的决策网络之前，往往需要经历数轮的开发迭代



## 小结：专家系统

- 一类具有专门知识和经验的计算机智能程序系统
  - 代表性的专家系统：DENDRAL、MYCIN、PROSPECTOR等
- 特点：启发性、透明性、灵活性
- 组成部分：知识库、推理机、数据库、知识获取机构、解释机构和人机接口
  - 应用案例：动物识别专家系统
- 决策专家系统：引入决策网络，基于信息价值推荐最优决策
  - 创建决策专家系统

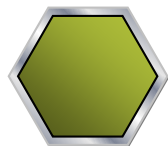
# 内容安排



效用理论



决策网络



信息价值



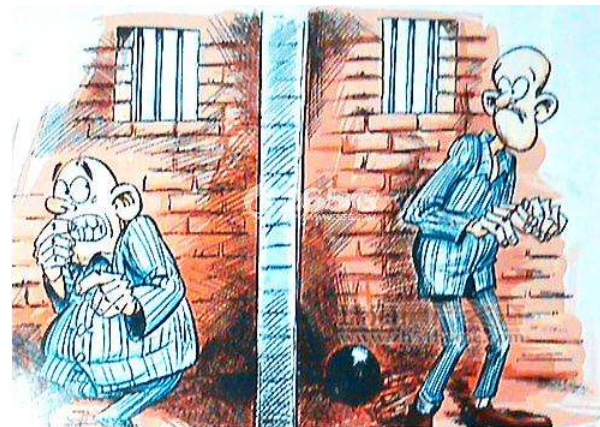
专家系统



单步博弈

# 囚徒困境

- 有两个囚徒（Agent）被隔离审讯
  - 一方可以揭发（Testify）另一方，也可以保持沉默（Refuse）
  - 如果A揭发B，B保持沉默，则A被释放，B判10年
  - 如果A和B相互揭发，则A和B各判5年
  - 如果A和B都保持沉默，则各判1年
- 假设两个Agent都知晓右图的效用矩阵
- 两个Agent需要在不知晓另一方行动的前提下，同时行动



		Agent 2	
		Testify	Refuse
Agent 1	Testify	-5, -5	0, -10
	Refuse	-10, 0	-1, -1

效用矩阵

# 策略

- 一个Agent的策略（strategy）分为：

- **纯策略**：行动的选择是确定性的
  - 如：以1的概率选择揭发
- **混合策略**：行动的选择是概率性的
  - 如：[揭发: 0.7; 沉默: 0.3]

纯策略是混合策略的一种特殊情形

- **混合策略的效用**可以写成纯策略的效用的线性组合形式：

$$U([a_1:p_1; \dots; a_n : p_n]) = \sum_{i=1}^n p_i U(a_i)$$

- $s_i$ ：Agent  $i$  的策略
- $s_{1:n}$ ：所有  $n$  个Agent的**策略组合**（strategy profile）
- $s_{-i}$ ：所有Agent中除去Agent  $i$  的策略组合

# 最优反应

- $U_i(s_{1:n})$  或  $U_i(s_i, s_{-i})$  : 给定一个策略组合  $s_{1:n}$  , Agent  $i$  的效用
- Agent  $i$  对策略组合  $s_{-i}$  的一个**最优反应** (best response) 是一个策略  $s_i^*$  , 满足:  
对所有策略  $s_i$  , 有  $U_i(s_i^*, s_{-i}) \geq U_i(s_i, s_{-i})$
- 一般地, 给定  $s_{-i}$  , 可以有多个不同的最优反应



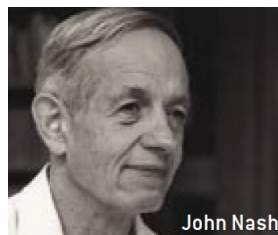
# 占优策略

- 在一些博弈中，存在一个 $s_i$ ，它是所有可能 $s_{-i}$ 的一个最优反应，称 $s_i$ 为一个占优策略（dominant strategy）
- 当所有Agent都使用占优策略，称它们的组合为占优策略均衡（dominant strategy equilibrium）
- 囚徒困境
  - 不论Agent 2揭发还是沉默，Agent 1揭发都会更好
  - 揭发是Agent 1的占优策略
  - 同理，揭发也是Agent 2的占优策略
  - 占优策略均衡：两个Agent均揭发对方，各判5年
  - 最好的结果：两个Agent均保持沉默，各判1年



当所有Agent均使用最优反应，得到的却是一个次优的结果！

# 纳什均衡



## ■ 避碰博弈

- 同时爬升（Climb）或下降（Descend）会增加碰撞危险，效用-4
- 爬升会消耗更多机油，额外效用-1
- 有两个纯策略的纳什均衡，即（Climb, Descend）和（Descend, Climb）

## ■ 不存在占优策略

- 一个飞行员的最优反应，依赖于另一个飞行员的行动

## ■ 如果对所有Agent $i$ ， $s_i$ 是 $s_{-i}$ 的一个最优反应，则 $s_{1:n}$ 是一个纳什均衡（Nash equilibrium）

- 单方面改变一个Agent的策略不能从中获益
- 可以证明，每个博弈至少有一个纳什均衡（不一定是纯策略）

		Agent 2	
		Climb	Descend
Agent 1	Climb	-5, -5	-1, 0
	Descend	0, -1	-4, -4

效用矩阵

# 旅行者困境

- 航空公司弄丢了两个旅行者（Lucy和Pete）的两个相同的箱子
- 让旅行者写下他们箱子的价值 $a_i$  和 $a_{-i}$ ，范围是\$2和\$100之间





# 旅行者困境的效用函数

## ■ 效用函数

$$U_i(a_i, a_{-i}) = \begin{cases} a_i & \text{if } a_i = a_{-i} \\ a_i + 2 & \text{if } a_i < a_{-i} \\ a_{-i} - 2 & \text{otherwise} \end{cases}$$

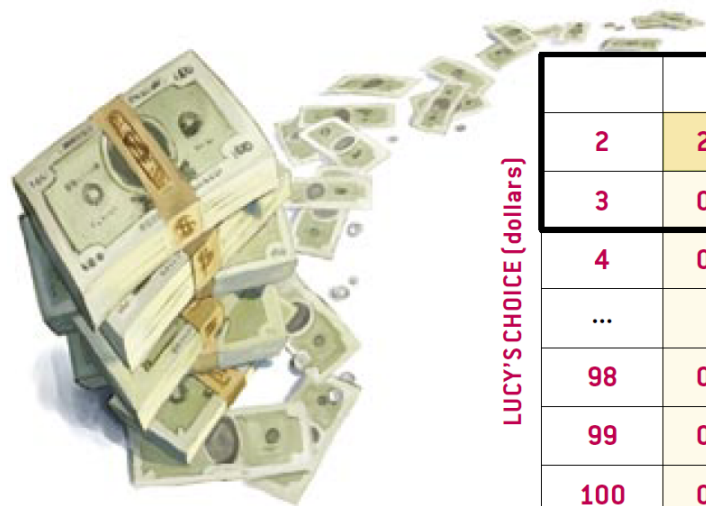
例1: Lucy报价\$100, Pete报价\$100  
赔偿给Lucy和Pete各\$100

例2: Lucy报价\$10, Pete报价\$90  
以\$10为基准,  
赔偿给Lucy: \$10 + \$2 = \$12  
赔偿给Pete: \$10 - \$2 = \$8

- 问: 大家倾向写多少钱?
- 大多数人倾向写\$97和\$100之间

# 旅行者困境的效用矩阵和纳什均衡

## ■ 效用矩阵



		PETE'S CHOICE (dollars)						
		2	3	4	...	98	99	100
LUCY'S CHOICE (dollars)	2	2 2	4 0	4 0	...	4 0	4 0	4 0
	3	0 4	3 3	5 1	...	5 1	5 1	5 1
	4	0 4	1 5	4 4	...	6 2	6 2	6 2
	...	...	...	...	...	...	...	...
	98	0 4	1 5	2 6	...	98 98	100 96	100 96
	99	0 4	1 5	2 6	...	96 100	99 99	101 97
	100	0 4	1 5	2 6	...	96 100	97 101	100 100

事实上，旅行者困境问题只有唯一一个纳什均衡：

(\$2, \$2)

# 行为博弈理论

- 当构建一个与人类博弈的决策系统时，人类经常不使用纳什均衡
  - 认知局限：计算困难
  - 存在多个纳什均衡
  - 怀疑对方是否遵循纳什均衡
- 行为博弈理论：建模人类Agent的行为
- 对率 $k$ 层模型（logit level- $k$  model），基于以下假设：
  - 当错误的成本越低时，犯这些错误的概率越高
  - 人类能够策略性向前推理的步数是有限的

我想你在想我在想...



# 捡帽子的故事



我想你会学  
我丢帽子

我想你在想我会  
学你丢帽子



# 对率 $k$ 层模型

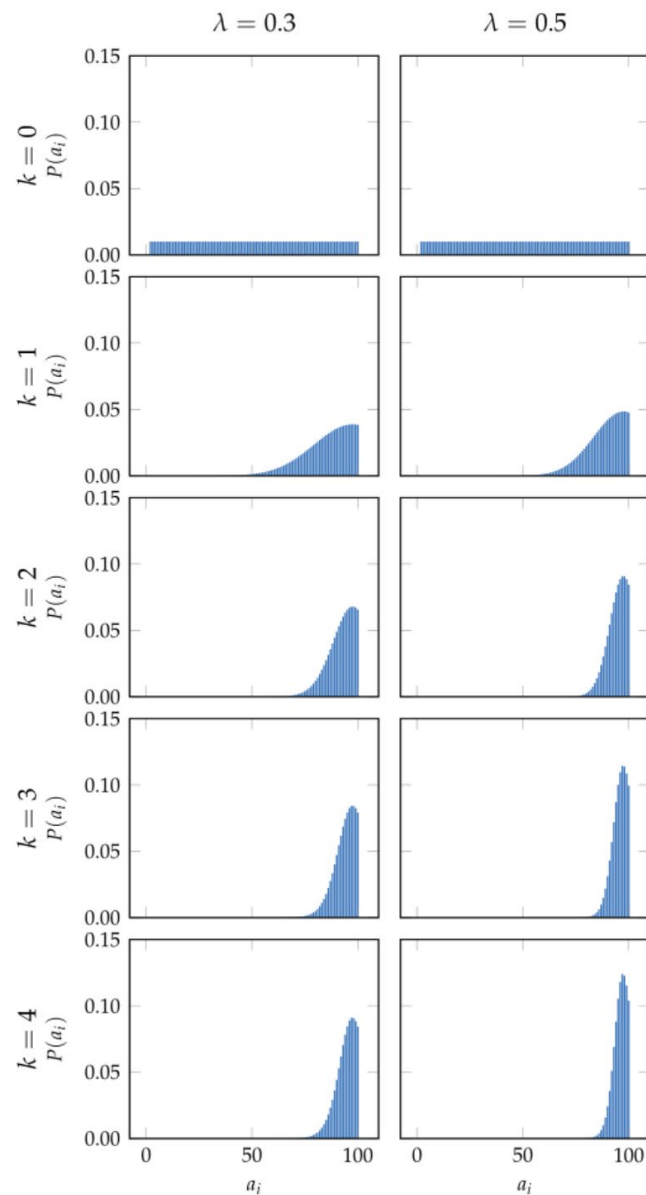
- 对率 $k$ 层模型的参数：从数据中学得
  - 准确（precision）参数 $\lambda \geq 0$ ：控制对效用差异的敏感度
  - 深度（depth）参数 $k > 0$ ：控制理性深度
- 第0层Agent：均匀地选择行动
- 第 $k$ 层Agent：假设对手采取的是第 $k - 1$ 层策略，根据对率分布来选择行动

$$P(a_i) \propto e^{\lambda U_i(a_i, s_{-i})}$$

$s_{-i}$ 表示其他Agent使用的假定策略

# 旅行者困境的对率 $k$ 层模型

- 右图：有不同 $\lambda$ 和 $k$ 值的对率 $k$ 层模型的策略
- 通常地，人类行为可以用对率2层模型很好建模
  - 在这个问题上，提供了比纳什均衡更好的人类行为模型



# 小结：单步博弈

## ■ 博弈

- 囚徒困境、避碰博弈、旅行者困境

## ■ 策略

- 纯策略、混合策略、策略组合
- 最优反应
- 占优策略
- 占优策略均衡

## ■ 纳什均衡

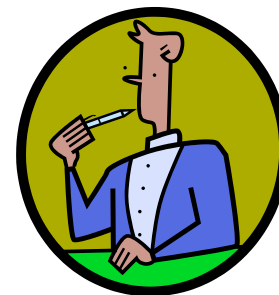
## ■ 行为博弈理论

- 对率 $k$ 层模型

## 课后练习3.3

- 解释信息价值。如果一个观察不改变最优行动，它的信息价值是多少？
- 证明信息的期望价值是非负的：

$$VOI(O_j | \mathbf{o}) \geq 0 \quad \forall \mathbf{o}, O_j$$





## 课后练习3.4

- 一个旧车购买者可以决定进行不同费用的各种测试（例如，踢轮胎，将车送到合格的汽车机械师处检查），然后，取决于这些测试的结果，决定购买哪辆车。我们将假设购车者正在考虑是否购买车 $c_1$ ，只有进行至多一次测试的时间； $t_1$ 是对 $c_1$ 的测试，费用\$50。

一辆车可以状况很好（质量为 $q^+$ ）或者状况很差（质量为 $q^-$ ），测试可能帮助指示该车所处的状况。购买车 $c_1$ 的费用为\$1500，如果它状况很好则它的市场价为\$2000；如果状况不好，需要花\$700来维修使它的状况变好。购车者的估计是，有70%的几率 $c_1$ 状况很好。

- （a）画出表示这个问题的决策网络。
- （b）不进行测试，计算购买 $c_1$ 的期望净获利。
- （c）给定车处于很好或者很差的状况，测试可以根据车通过还是不通过该测试的概率进行描述。我们有下列信息：

$$P(\text{pass}(c_1, t_1) | q^+(c_1)) = 0.8$$

$$P(\text{pass}(c_1, t_1) | q^-(c_1)) = 0.35$$

计算车通过（或者通不过）测试的概率，并使用贝叶斯规则计算出在给定每个可能的测试结果条件下，车处于好（或者不好）的状况的概率。

- （d）给定通过或者通不过测试，分别计算买车和不买车的期望效用，并用最大化期望效用原则分析出每种测试结果对应的最优行动。

- （e）计算测试的信息价值，并且为购车者产生一个最优条件规划。



## 课后练习3.5

- 对囚徒困境做如下修改：
  - 如果A揭发B，B保持沉默，则A被释放，B判4年
  - 如果B揭发A，A保持沉默，则B被释放，A判4年

问：修改后的博弈还有占优策略均衡吗？还有哪些纳什均衡？

交第一次课后作业（第1~3部分）的截止时间为：

**2021年4月21日**

