南京大学 人工智能学院
School of Artificial Intelligence, Nanjing University

第三部分：简单决策系统

章宗长

2020年3月27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

效用理论

- **偏好与效用**
- 期望效用最大化
- 效用函数
- 多变量效用函数
- 人类评价与非理性

偏好

- 信念度：比较两个不同陈述（事件）的可信程度
- 偏好：比较两种不同结果的渴求程度
- 用如下记号描述一个Agent的偏好：
  - $A \succ B$　Agent偏好A甚于B
  - $A \sim B$　Agent对A和B偏好相同
  - $A \succcurlyeq B$　Agent偏好A甚于B或者偏好相同
- 与信念度一样，偏好也是主观的
- 彩票抽奖：每个行动为抽一张彩票，可能结果为$S_{1:n}$，其发生概率分别为$p_{1:n}$的一次抽奖记为：

$$[S_1 : p_1; \ldots; S_n : p_n]$$

理性偏好的约束

- **完整性：** $A \succ B$， $B \succ A$或$A \sim B$中必有一个成立
- **传递性：** 如果$A \succcurlyeq B$和$B \succcurlyeq C$，则$A \succcurlyeq C$
- **连续性：** 如果$A \succcurlyeq C \succcurlyeq B$，则存在概率$p$使得

$$[A:p;\ B:1-p] \sim C$$

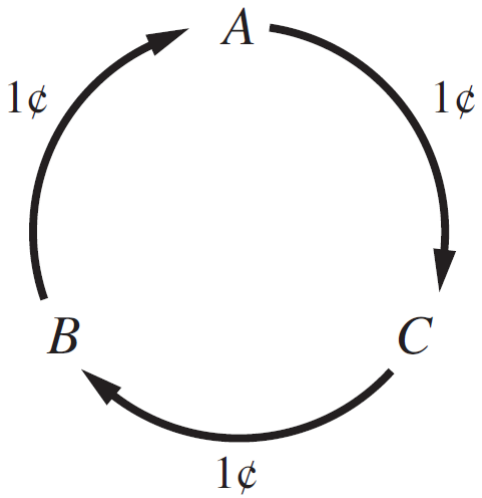
- **独立性：** 如果$A \succ B$，则对于任何C和概率p，有

$$[A:p;\ C:1-p] \succcurlyeq [B:p;\ C:1-p]$$

- 这些理性偏好遵循的约束被称为：冯·诺依曼-摩根斯坦公理
- 违背任一约束将在某些情况下展现出明显不理性的行为

理性偏好的约束（续）

- 例子：非传递性偏好$A \succ B \succ C \succ A$导致了不理性行动



可以自由交换的商品：A, B, C

$C = A + 1¢$

$B = C + 1¢$

$A = B + 1¢$

具有非传递性偏好的Agent会掏出所有钱

- 人有时不是很理性的

- **我们的目标**：从计算的角度理解理性决策，以便构建有用的系统，而不是理解人类如何做决策的

偏好导致效用

- 由理性偏好的约束可推导出，存在一个实数效用函数U，使得：
  - $U(A) > U(B)$ 当且仅当 $A \succ B$
  - $U(A) = U(B)$ 当且仅当 $A \sim B$
- 如果一个Agent的效用函数根据如下公式进行变换，它的行为将不会改变：

$$U'(s) = mU(s) + b$$ 其中m和b是常数， $m > 0$

仿射变换

- 效用好像温度，可以用开尔文、华氏、摄氏等度量系统比较不同的温度，这些度量可以由彼此的仿射变换得到

偏好导致效用（续）

- 由理性偏好的约束可得，一次抽奖的效用：

$$U([S_1:p_1;\ldots;S_n:p_n]) = \sum_{i=1}^{n} p_i U(S_i)$$

- 假设构建一个避碰系统，与一架飞机相遇的结果由系统是否发出警报（A）和碰撞是否发生（C）来定义：
  - A和C是二进制的，有4种可能的结果
  - 只要偏好是理性的，可以定义在这些结果上的效用：
    $U(a^0,c^0)$，$U(a^1,c^0)$，$U(a^0,c^1)$，$U(a^1,c^1)$

- 将其发生概率设为$p_{1:4}$，有

$$U([a^0,c^0:p_1;a^1,c^0:p_2;a^0,c^1:p_3;a^1,c^1:p_4])$$

等价于

$$p_1U(a^0,c^0)+p_2U(a^1,c^0)+p_3U(a^0,c^1)+p_4U(a^1,c^1)$$

效用理论

- 偏好与效用
- **期望效用最大化**
- 效用函数
- 多变量效用函数
- 人类评价与非理性

最大化期望效用原则

- 如何在环境状态不完全可观察时，做**理性决策**？
- 假设有一个概率模型$P(s'|a,o)$：Agent采取了行动a、得到观察o，环境变为s'的概率
- 给定观察o，采取行动a的**期望效用**：

$$EU(a|o) = \sum_{s'} P(s'|a,o)\,U(s')$$

- **最大化期望效用原则**（Maximum Expected Utility, MEU）：理性Agent应该采取能最大化期望效用的行动

$$a^* = \arg\max_a EU(a|o)$$

- 我们感兴趣的是理性Agent，MEU是智能系统设计的**中心原则**

期望效用与后决策失望

- 在很多时候，我们实际使用的是真实期望效用的估计值

$$\widehat{EU}(a \mid o)$$

- 假设$\widehat{EU}(a \mid o)$是$EU(a \mid o)$的无偏估计：

$$E(\widehat{EU}(a \mid o) - EU(a \mid o)) = 0$$

？

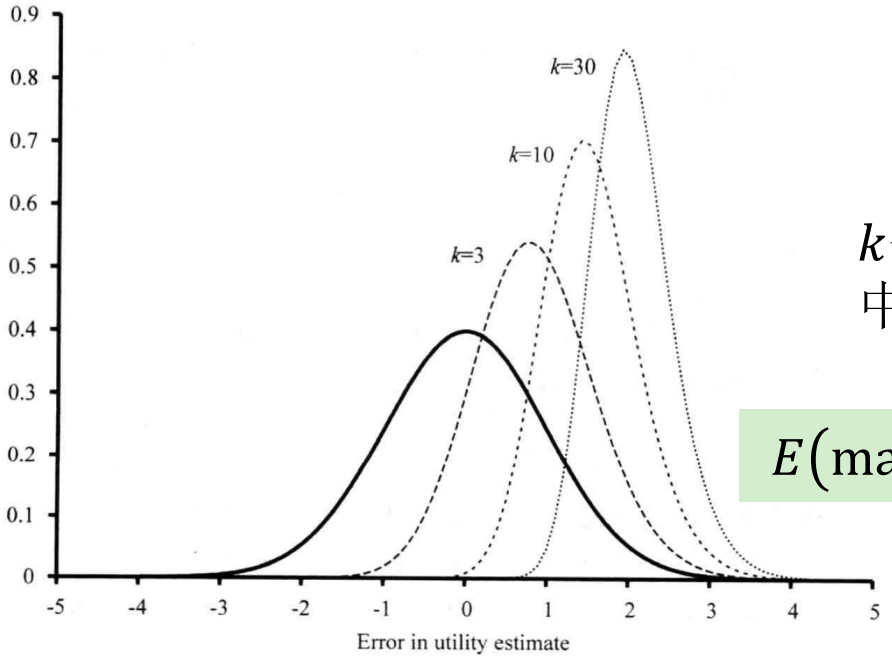
$$E(\max_a \widehat{EU}(a \mid o) - \max_a EU(a \mid o)) = 0$$

- 即使估计值是无偏的，真实结果通常比估计的要差很多：

$$E(\max_a \widehat{EU}(a \mid o) - \max_a EU(a \mid o)) \geq 0$$

期望效用与后决策失望（续）

- 考虑一个决策问题：有k个选项，每一个的真实效用是0。假设每个效用估计值的误差具有0均值，1标准差



k个效用估计的误差，以及k个估计中最大值的分布（$k = 3$，10，30）

$$E\left(\max_a \widehat{EU}(a \mid o) - \max_a EU(a \mid o)\right) > 0$$

- **乐观者报应**：最优选项的估计期望效用会高于真实值的现象

效用理论

- 偏好与效用
- 期望效用最大化
- **效用函数**
- 多变量效用函数
- 人类评价与非理性

效用的获取

- 在许多情境中，获得一个合适的效用函数比获得一个概率模型更加困难
- 概率通常可以从数据中学习，或者从专家那里获得
- 效用因人而异，不能被直接观察到

效用启发式

- 效用启发式（偏好启发式）：根据人的经验，推导出Agent的效用函数的可能形式
- 归一化的效用函数：最好结果的效用为1，最坏结果的效用为0，其他结果的效用介于0和1之间
- 一种方法：通过固定两个特殊结果的效用来建立一个尺度
  - 如：固定水的结冰点和沸点来建立温度的尺度
  - 把最坏结果$S_{\perp}$固定为0，最好结果S_{T}固定为1
  - 对结果S，调节概率p直到Agent对S和标准抽奖$[S_{\mathrm{T}}: p; S_{\perp}: 1-p]$**没有偏向性**。在归一化效用下，$S$的效用是$p$

效用启发式（续）

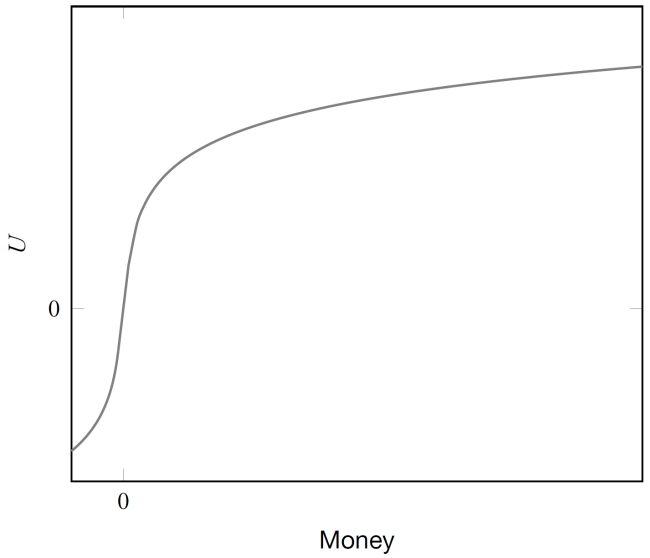
避碰示例

- 最好的结果：没有报警也没有碰撞，即$U(a^0, c^0) = 1$
- 最坏的结果：有报警也有碰撞，即$U(a^1, c^1) = 0$
- 定义抽奖：$L(p) = [a^0, c^0 : p; a^1, c^1 : 1 - p]$
- 为了确定$U(a^1, c^0)$，需要找到p，使得$(a^1, c^0) \sim L(p)$
- 类似地，为了确定$U(a^0, c^1)$，需要找到p，使得$(a^0, c^1) \sim L(p)$

人类生命的效用

- 在医疗、交通等决策问题中，人们的生命面临危险
  - 为人的生命设置一个效用值
- 错误的假设：人的效用与死亡的概率是线性关系，即某个死亡概率p的结果的效用估计为$p \cdot U(\text{死亡})$
- 微亡（micromort）：百万分之一的死亡风险
- 质量调整寿命（Quality-Adjusted Life Years, QALY）
  - 健康无疾病的生命的一年的值为1QALY
  - 有病的生命的一年的值低于1QALY

货币效用



- 在经济学里，货币效用关于货币总量通常是非线性的
- 当数额不大时，货币的效用曲线大体是线性的
  - 如：$100的效用是$50的两倍
- 当数额很大时，货币的效用曲线大体通常是对数的
  - 如：对于亿万富翁，$1000的作用不像对于普通人群那么大
- 保险购买策略的期望货币价值总是负的

圣彼得堡悖论

- 假定你有机会玩一个游戏，其中一枚无偏的硬币被重复投掷直到正面朝上。假如第一次正面朝上出现在第n次投掷时，你可以获得$\$2^n$。那么，你愿意付多少钱来获得玩这样一次游戏的机会？
- 大部分人只愿意付$\$2$
- 事件H_n：首次正面朝上出现在第n次投掷时
  - $P(H_n) = 1/2^n$
- 期望收益：

如果把货币总量作为效用函数，则你应该愿意付任意多的钱来玩这个游戏

$$\sum_{n=1}^{\infty} P(H_n)\text{Payoff}(H_n) = \sum_{n=1}^{\infty} (1/2^n)2^n = 1 + 1 + \cdots = \infty$$

圣彼得堡悖论（续）

- 如果采用

$$U\big(\text{Payoff}(H_n)\big) = \log_2 \text{Payoff}(H_n),$$

那么可以得到

$$\sum_{n=1}^{\infty} P(H_n) U\big(\text{Payoff}(H_n)\big) = \sum_{n=1}^{\infty} (1/2^n) \log_2 2^n = 2$$

这恰恰是大部分人玩这个游戏所愿付出的钱的数量

- 经验心理学的研究表明，$U(k) = \alpha + \beta \log(k + \gamma)$

风险态度

假设A：得到50元；B：有50%概率得到100元

- 风险中立：效用函数是线性的
  - Agent对A和B的偏好相同（$A \sim B$）
- 追求风险：效用函数是朝上凹的
  - Agent对B的偏好甚于A（$B \succ A$）
- 规避风险：效用函数是朝下凹的
  - Agent对A的偏好甚于B（$A \succ B$）

效用理论

- 偏好与效用
- 期望效用最大化
- 效用函数
- **多变量效用函数**
- 人类评价与非理性

多变量效用函数

- 令有n个变量的效用函数为$U(x_{1:n})$

- 例子：避碰系统的效用函数依赖于两个二值变量
  - 是否有警报和是否有碰撞
  - 需要指定在4种可能组合上的效用

- 例子：在避碰系统中加入两个额外的变量
  - 强化（S）：是否加强爬升或下降
  - 逆转（R）：是否改变飞行方向
  - 需要指定在2^4 =16种可能组合上的效用

多变量效用函数（续）

- 例子：依赖于n个二值变量的效用函数
  - 需要指定在2^n种可能组合上的效用
  - 若归一化该函数，则至少有一个值为0，至少有一个值为1

- 通过利用变量间不同形式的**独立性**，可以压缩表示效用函数：

$$U(x_1, \cdots, x_n) = f[f_1(x_1), \cdots, f_n(x_n)]$$

偏好独立性

- 令$\boldsymbol{X}$, $\boldsymbol{Y}$是效用变量集$\boldsymbol{V}$的不相交划分，$\boldsymbol{X}$在$\succ$上偏好独立于$\boldsymbol{Y} = \boldsymbol{V} - \boldsymbol{X}$，如果对于所有的$\boldsymbol{y}, \boldsymbol{y}' \in Val(\boldsymbol{Y})$以及所有的$\boldsymbol{x}_1, \boldsymbol{x}_2 \in Val(\boldsymbol{X})$，有

$$\boldsymbol{x}_1 \succ_{\boldsymbol{y}} \boldsymbol{x}_2 \Leftrightarrow \boldsymbol{x}_1 \succ_{\boldsymbol{y}'} \boldsymbol{x}_2$$

- 例子：如果结果$\langle x_1, x_2, x_3 \rangle$和$\langle x_1', x_2', x_3 \rangle$之间的偏好不依赖于变量$X_3$的任一具体值$x_3$，则称两个变量$X_1$和$X_2$偏好独立于第三个变量$X_3$
- 如果效用变量集$\boldsymbol{V}$中的任一变量在$\succ$上都偏好独立于其补集，则称变量集$\boldsymbol{V}$满足相互偏好独立性

偏好独立性（续）

- 例子：考虑一名企业家，其效用函数$U(S,F)$涉及两个二值属性：其公司取得成功（S）和个人出名（F）。结果上的一个理性偏好次序是：

$$(s^1,f^1) \succ (s^1,f^0) \succ (s^0,f^0) \succ (s^0,f^1)$$

- S偏好独立于F，因为

$$(s^1,f^1) \succ (s^0,f^1), (s^1,f^0) \succ (s^0,f^0)$$

- F并不偏好独立于S，因为

$$(s^1,f^1) \succ (s^1,f^0), (s^0,f^1) \prec (s^0,f^0)$$

- 偏好独立性并不是对称的关系

加法效用函数

- 如果变量$X_{1:n}$是相互偏好独立的，那么可以使用单一变量效用函数之和来表示一个多变量效用函数：

$$U(x_{1:n}) = \sum_{i=1}^{n} U(x_i)$$

加法效用函数

- 假设所有变量是二值的，则仅需$2n$个值来指定该效用函数：

$$U(x_1^0),\ U(x_1^1),\ \ldots,\ U(x_n^0),\ U(x_n^1)$$

例子：避碰系统

- 具有4个变量的避碰系统：利用相互偏好独立性假设，仅需8个值来指定效用函数

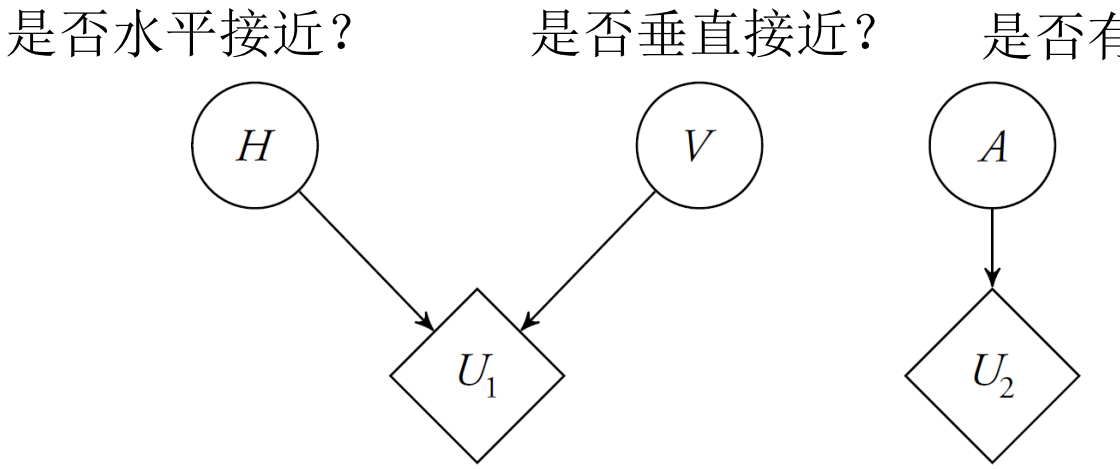
效用启发式

- 如果没有报警、碰撞、强化、逆转，则对应单一变量的效用$U(a^0)$、$U(c^0)$、$U(s^0)$和$U(r^0)$为0
  - 从而有$U(a^0, c^0, s^0, r^0) = 0$
- 碰撞的成本最高，设置$U(c^1) = 1$
- 这样下来，只需要3个值来定义效用函数

效用函数的分解

- 很多问题的效用函数不能写成单一变量效用函数的加法分解形式
- 例子：用3个二值变量来定义避碰系统的效用函数
  - 是否入侵者在水平方向接近（H）
  - 是否入侵者在垂直方向接近（V）
  - 是否系统发出警报（A）
- 仅当同时有h^1和v^1，碰撞的威胁才真正存在
- H和V之间不满足互相偏好独立性假设，因此不能对所有变量使用加法分解
- 但是，可以有$U(h, v, a) = U(h, v) + U(a)$

效用函数的分解（续）



效用函数的分解示意图

- **效用结点**（菱形）
- **不确定性结点**（圆形）
  - 效用结点的父结点：效用结点所依赖的结点
  - 离散：效用函数可用表格表示
  - 连续：可用任一实数函数来表示效用函数
- 如果有多个效用结点，则总效用值为这些效用结点的值之和

效用独立性

- 令$\boldsymbol{X}, \boldsymbol{Y}$是效用变量集$\boldsymbol{V}$的不相交划分，$\boldsymbol{X}$在$\succ$上效用独立于$\boldsymbol{Y} = \boldsymbol{V} - \boldsymbol{X}$，如果对于所有的$\boldsymbol{y}, \boldsymbol{y}' \in Val(\boldsymbol{Y})$，以及$Val(\boldsymbol{X})$上的任意一对抽奖$L_1, L_2$，有

$$L_1 \succ_{\boldsymbol{y}} L_2 \Leftrightarrow L_1 \succ_{\boldsymbol{y}'} L_2$$

- 命题：集合$\boldsymbol{X}$在$\succ$上效用独立于$\boldsymbol{Y} = \boldsymbol{V} - \boldsymbol{X}$，当且仅当$U(\boldsymbol{V}) = f(\boldsymbol{Y}) + g(\boldsymbol{Y})h(\boldsymbol{X})$

- 如果效用变量集$\boldsymbol{V}$中的任一变量在$\succ$上都效用独立于其补集，则称变量集$\boldsymbol{V}$满足相互效用独立性

乘法效用函数

- 如果变量集$\boldsymbol{V}$满足相互效用独立性，则效用函数$U(\boldsymbol{V})$可以表示为一个乘法效用函数

- 例子：令U_i表示$U(x_i)$，如果$\boldsymbol{X} = \{x_1, x_2, x_3\}$满足相互效用独立性，则

$$U(x_1, x_2, x_3) = k_1U_1 + k_2U_2 + k_3U_3 + k_1k_2U_1U_2 + k_2k_3U_2U_3 + k_1k_3U_1U_3 + k_1k_2k_3U_1U_2U_3$$

- 呈现相互效用独立性的n变量问题可以用n个单一变量效用函数和n个常数来建模

效用理论

- 偏好与效用
- 期望效用最大化
- 效用函数
- 多变量效用函数
- **人类评价与非理性**

人类评价与非理性

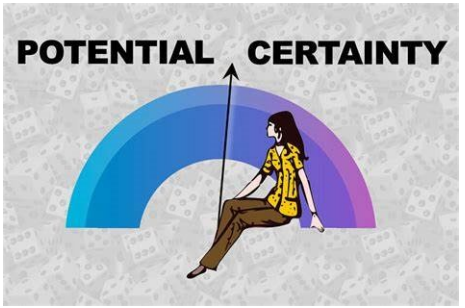
- 决策理论：一种规范性理论，描述了一个理性的Agent应该如何行动
- 描述性理论：描述了实际的Agent（例如人类）真正如何行动
- 有一些实验证据表明这两者不是一致的
- 人类的偏好在很多时候是非理性的

Allais悖论

- 在两次抽奖A和B之间选择
  - A：80%的机会获得$4000
  - B：100%的机会获得$3000
- 在C和D之间选择
  - C：20%的机会获得$4000
  - D：25%的机会获得$3000
- 大部分人偏好选择B而不选择A，选择C而不选择D
  - $B \succ A$蕴含着$U(\$3000) > 0.8U(\$4000)$
  - $C \succ D$蕴含着$0.2U(\$4000) > 0.25U(\$3000)$

$$0.8U(\$4000) > U(\$3000)$$

- 没有效用函数能够与这些选择一致
- 非理性偏好的一个解释：**确定性效应**
  - 人们被确定性的收益高度吸引



Ellsberg悖论

- 缸里面有1/3的球是红色的，剩下2/3的球是黑色或黄色的，但不知道有多少黑球和多少黄球

- 选择A或B作为奖励规则
  - A：取得红球得$100
  - B：取得黑球得$100

- 选择C或D作为奖励规则
  - C：取得红球或黄球得$100
  - D：取得黑球或黄球得$100

- 大部分人偏好选择A而不选择B，选择D而不选择C

- 大多数人选择已知的概率，而不愿意选择未知的东西

表达效应

- 表达一：一个医疗过程有90%的生还率
- 表达二：一个医疗过程有10%的死亡率

- 人们对前者的喜欢程度大约是后者的两倍

- 表达效应（framing effect）：一个决策问题的措辞对Agent的选择有很大的影响

锚效应

- 锚效应（anchoring effect）：人们对进行相对效用评价感觉更舒服，而不愿意进行绝对的评价

- 例子：服装店提供各式各样的衣服，利用**锚效应**，在醒目的位置摆$1000的衣服，这会使顾客对所有衣服的价格估计偏高，最后买了$200的衣服，感觉很便宜

小结：效用理论

- 效用函数
  - 偏好、理性偏好的约束（冯·诺依曼-摩根斯坦公理）
  - 效用启发式
  - 货币效用：非线性、圣彼得堡悖论
  - 三种类型的Agent（风险中立、追求风险、规避风险）
  - 多变量效用函数：偏好独立性、效用独立性、效用函数分解
- 最大化期望效用原则
  - 期望效用、后决策失望（乐观者报应）
- 人类评价与非理性
  - Allais悖论、Ellsberg悖论、表达效应、锚效应

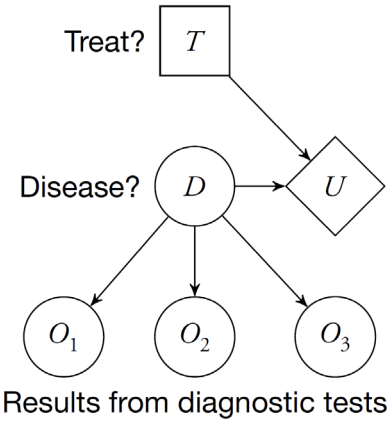
内容安排

- 效用理论
- **决策网络**
- 信息价值
- 专家系统
- 单步博弈

决策网络

- 决策网络（影响图）：贝叶斯网络+行动、效用
- 由三种类型的结点构成
  - 机会结点（椭圆）：随机变量
  - 决策结点（矩形）：在该结点上决策制定者有一个对行动的选择
  - 效用结点（菱形）：Agent的效用函数

机会结点：D, O_1, O_2, O_3
决策结点：T
效用结点：U



诊断测试的决策网络

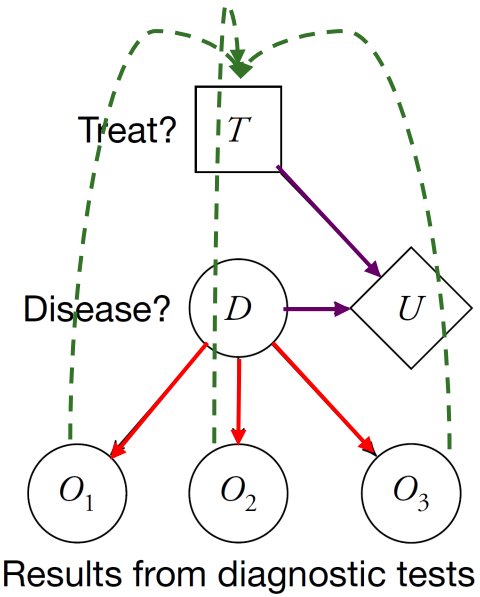
T	D	$U(T,D)$
0	0	0
0	1	−10
1	0	−1
1	1	−1

诊断测试的效用函数

决策网络（续）

三种类型的有向边

- **条件边**：指向机会结点，表明机会结点的不确定性条件于其所有父结点的值
- **信息边**：指向决策结点，表明该结点的决策由其父结点的值决定
  - 用虚线表示，有时省略
- **功能边**：指向效用结点，表明效用结点由其父结点的值决定



诊断测试的决策网络

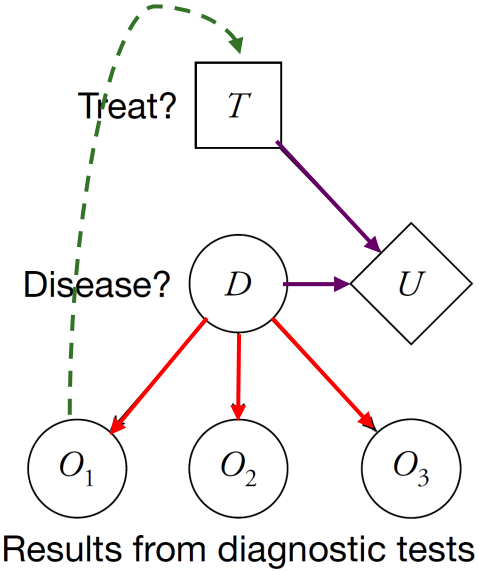
- **把决策问题表示为决策网络**
  - 利用问题的结构来计算基于效用函数的最优决策

评价决策网络

- 给定观察o，采取行动a的期望效用：

$$EU(a \mid o) = \sum_{s'} P(s' \mid a, o)\, U(s')$$

s'表示决策网络中结点的实例



- **计算治疗一种疾病的期望效用**

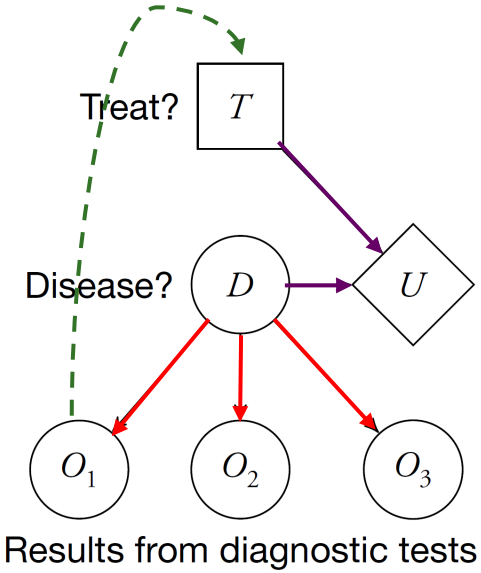
- 假设仅有第一次诊断测试的结果（正面的，记为o_1^1）
  - 添加一条O_1从到T的信息边，有

可以用贝叶斯网络的链式规则和条件概率的定义计算

$$EU(t^1 \mid o_1^1) = \sum_{o_3} \sum_{o_2} \sum_{d} P(d, o_2, o_3 \mid t^1, o_1^1)\, U(t^1, d, o_1^1, o_2, o_3)$$

评价决策网络（续）

- 效用结点仅依赖于是否出现了疾病和我们是否治疗它，因此可以把$U(t^1, d, o_1^1, o_2, o_3)$简化为$U(t^1, d)$



- 从而

$$EU(t^1 \mid o_1^1) = \sum_d P(d \mid t^1, o_1^1)\, U(t^1, d)$$

- 用前面介绍的任一精确、近似推理方法来评估$P(d \mid t^1, o_1^1)$
- 计算$EU(t^1 \mid o_1^1)$和$EU(t^0 \mid o_1^1)$
  - 若$EU(t^1 \mid o_1^1) > EU(t^0 \mid o_1^1)$，则最优行动为$t^1$
  - 若$EU(t^1 \mid o_1^1) = EU(t^0 \mid o_1^1)$，则最优行动为$t^0$或者$t^1$
  - 否则，最优行动为t^0

评价决策网络的算法

（1）把观察到的机会结点实例化为证据变量

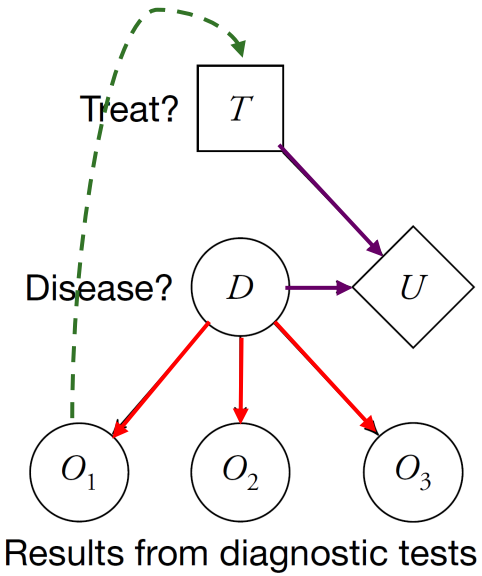
（2）对于决策结点的每个值：

（a）把决策结点设为该值

（b）对效用结点的父结点，使用一个概率推理算法计算其后验概率

（c）为该行动计算结果效用

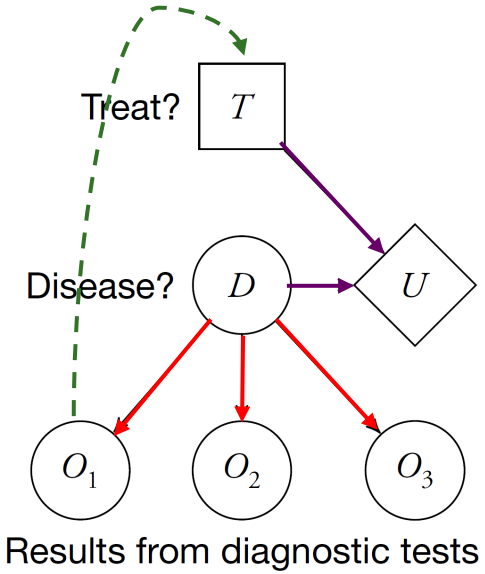
（3）返回最高效用的行动



- **改进算法：**如果行动结点和机会结点在决策网络中没有由（条件、信息、功能）边定义的孩子结点，则将它们移除
  - 右上图中，可以移除O_2和O_3，但不能移除O_1

小结：决策网络

- 决策网络
  - 贝叶斯网络+行动、效用
  - 三种类型的结点
    - 机会结点（椭圆）
    - 决策结点（矩形）
    - 效用结点（菱形）
  - 三种类型的有向边
    - 条件边（指向机会结点）
    - 信息边（指向决策结点）
    - 功能边（指向效用结点）
- 评价决策网络
  - 示例、算法及改进



诊断测试的决策网络

课后练习3.1

- Chris考虑从4辆二手车中买一辆具有最大期望效用的车。Pat考虑从10辆车中做同样的事。其他条件是相同的，两个人中哪个更可能买到更好的车？哪一个更可能会对其买到的车的质量失望？

课后练习3.2

- 考虑一位学生，他可以选择买或不买某门课程的教材。我们将用决策问题来建模，它有一个布尔决策结点B（指示Agent是否选择购买教材），和两个布尔机会结点M（指示该学生是否掌握了教材的内容）和P（指示该学生是否通过了考试）。当然，还有一个效用结点U。某个学生Sam有一个加法效用函数：不购买教材是0，购买是$-\$100$；通过考试是$\$2000$，没有通过是0。Sam的条件概率估计如下：

$$P(p \mid b, m) = 0.9 \qquad P(m \mid b) = 0.9$$
$$P(p \mid b, \neg m) = 0.5 \qquad P(m \mid \neg b) = 0.7$$
$$P(p \mid \neg b, m) = 0.8$$
$$P(p \mid \neg b, \neg m) = 0.3$$

你也许认为给定M下P是独立于B的，但这门课最后是开卷考试，所以有教材可能是有帮助的。

（1）画出该问题的决策网络。

（2）计算出购买和不购买教材的期望效用。

（3）Sam应该如何做？

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

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

$$VOI(O' \mid \boldsymbol{o}) = \left(\sum_{o'} P(o' \mid \boldsymbol{o})\, EU^*(\boldsymbol{o}, o') \right) - EU^*(\boldsymbol{o})$$

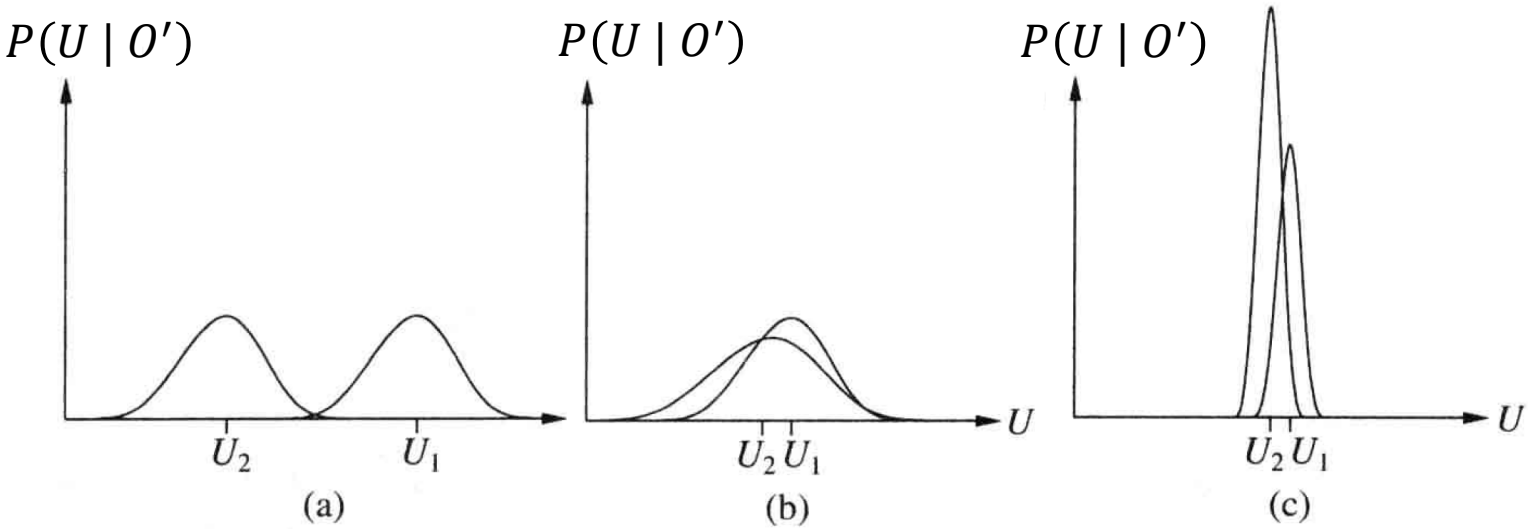
获取变量O'的观察之后的最优行动的期望效用

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，$VOI(O' \mid \boldsymbol{o})$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：

$$VOI(O_j \mid \boldsymbol{o}) \geq 0 \quad \forall \boldsymbol{o}, O_j$$

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：

$$VOI(O_i, O_j \mid \boldsymbol{o}) \neq VOI(O_i \mid \boldsymbol{o}) + VOI(O_j \mid \boldsymbol{o})$$ 一般情况下

- 信息价值是独立于次序的：

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

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

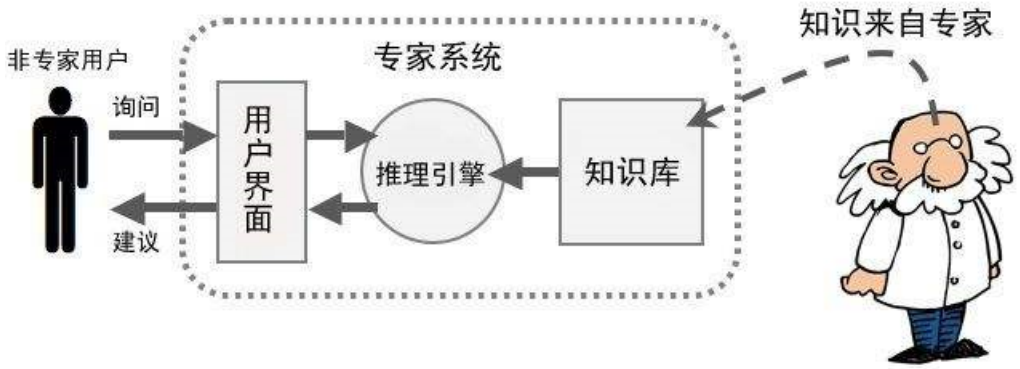
专家系统

- 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

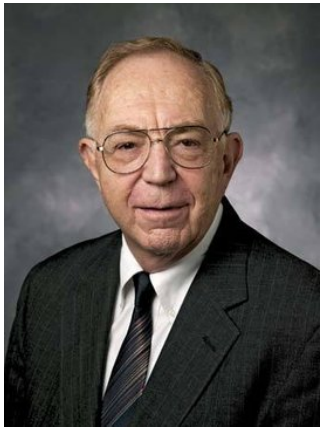
专家系统=知识库+推理机



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



费根鲍姆

勒德贝格

- 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

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

专家系统的类型

- 解释专家系统：通过对已有信息和数据的分析与解释，确定它们的涵义
  - 例子：化学结构分析、地质勘探数据解释
- 预测专家系统：通过对过去和现在已知状况的分析，推断未来可能发生的情况
  - 例子：恶劣气候预报、农作物病虫害预报
- 诊断专家系统：根据观察到的情况来推断出某个对象机能失常（即故障）的原因
  - 例子：肝功能检验、青光眼治疗

专家系统的类型（续）

- 设计专家系统：根据设计要求，求出满足设计问题约束的目标配置
  - 例子：大规模集成电路设计、齿轮加工工艺设计
- 规划专家系统：寻找出某个能够达到给定目标的动作序列或步骤
  - 例子：军事指挥调度、机器人规划
- 监视专家系统：不断观察系统的行为，并把观察到的行为与其应有的行为进行比较，以发现异常情况，发出警报
  - 例子：核电站的安全监视、传染病疫情监视

专家系统的类型（续）

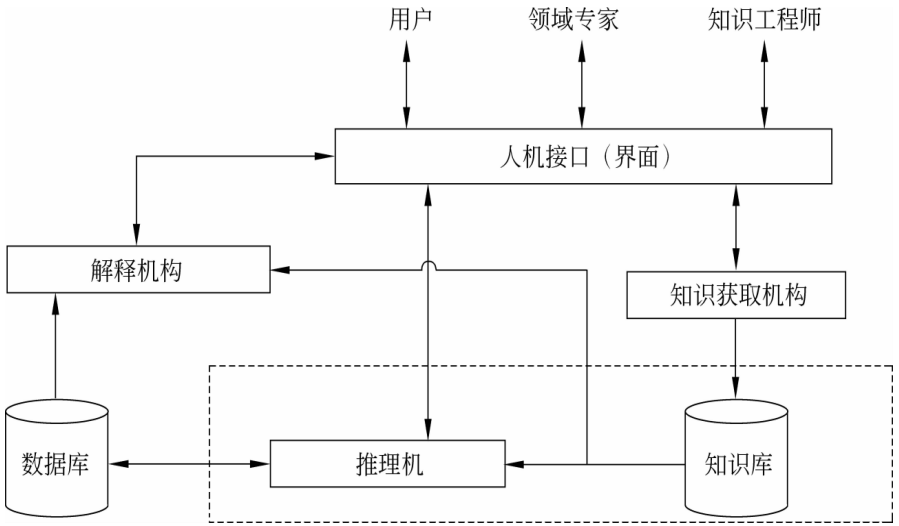
- 控制专家系统：自适应地管理一个受控对象或客体的全面行为，使之满足期望要求
  - 例子：空中交通管制、生产质量控制
- 教学专家系统：根据学生的特点、弱点和基础知识，以最适当的教案和教学方法对学生进行教学和辅导
  - 例子：计算机程序设计语言辅助教学、聋哑人语言训练
- 决策专家系统：引入决策网络，推荐能反映出Agent的偏好和可得到的证据的最优决策
  - 例子：心脏病的医疗方案推荐、飞行避碰方案推荐

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

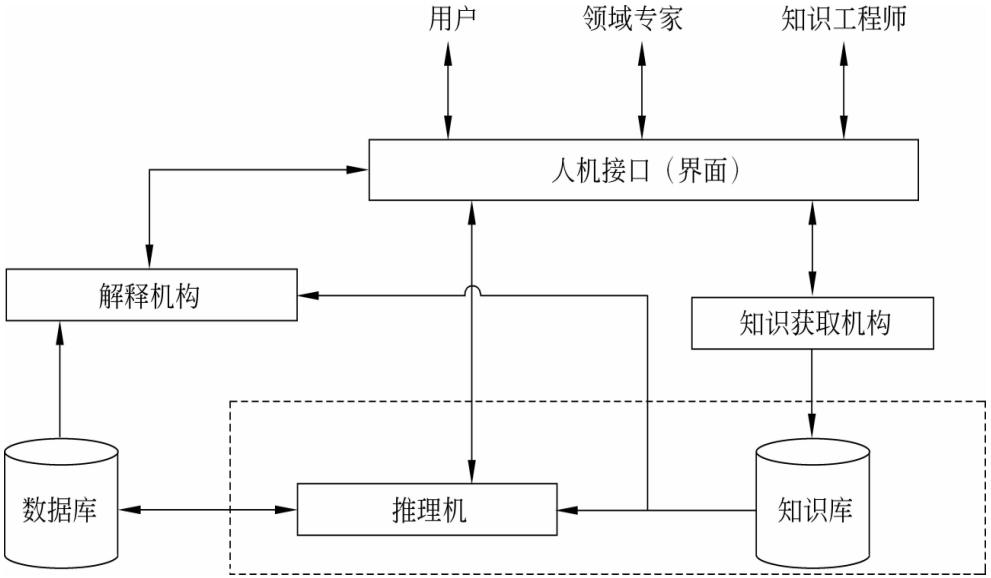


- 知识库：问题求解所需要的领域知识的集合
- 推理机：问题求解的核心执行机构

- 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

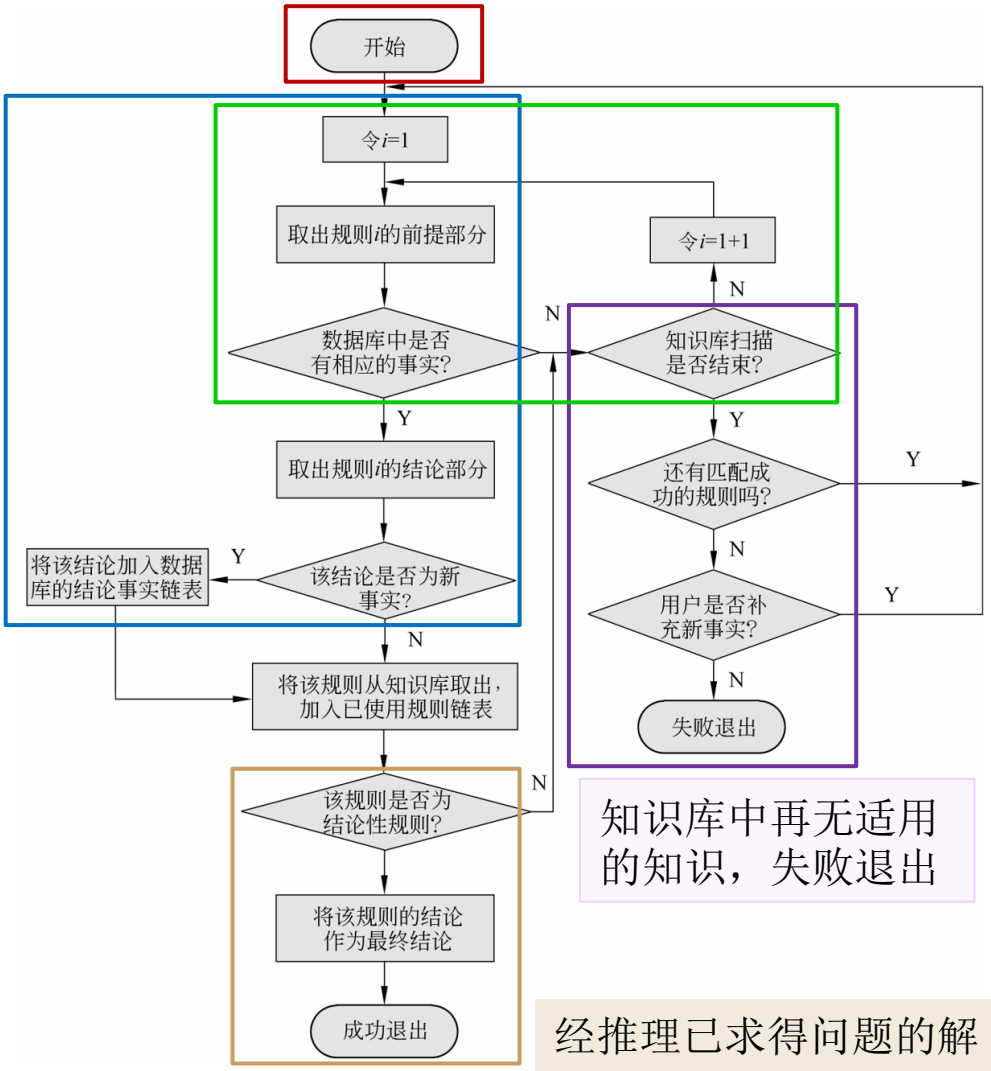
应用案例：动物识别专家系统（续）

- 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、驼鸟、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%

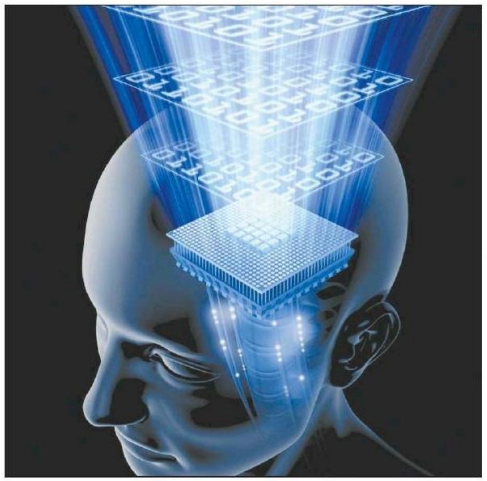
这可能是灾难性的！

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联的机会结点、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 1 \ Agent 2	Testify	Refuse
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：

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

- 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$U_i(s_i^*, s_{-i}) \geq U_i(s_i, s_{-i})$

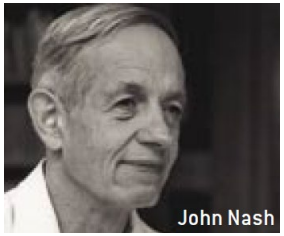
- 一般地，给定s_{-i}，可以有多个不同的最优反应

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）

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

效用矩阵

- 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）

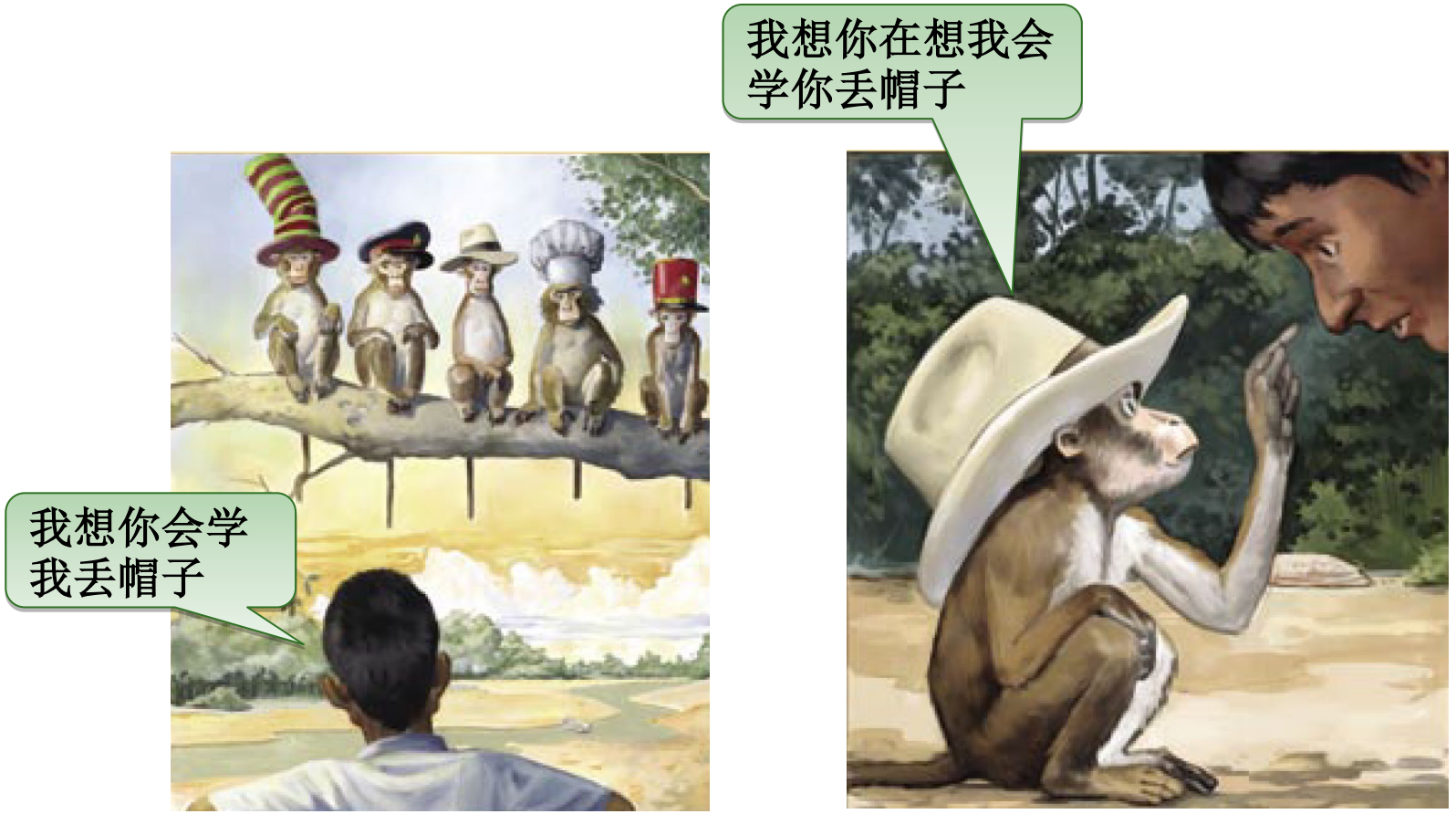
行为博弈理论

- 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

捡帽子的故事



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

- 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

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

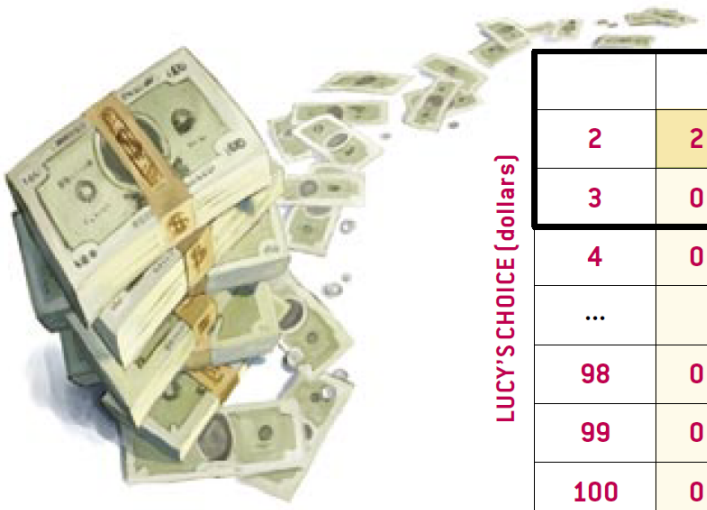
例1：Lucy报价\$100，Pete报价\$100
赔偿给Lucy和Pete各\$100

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



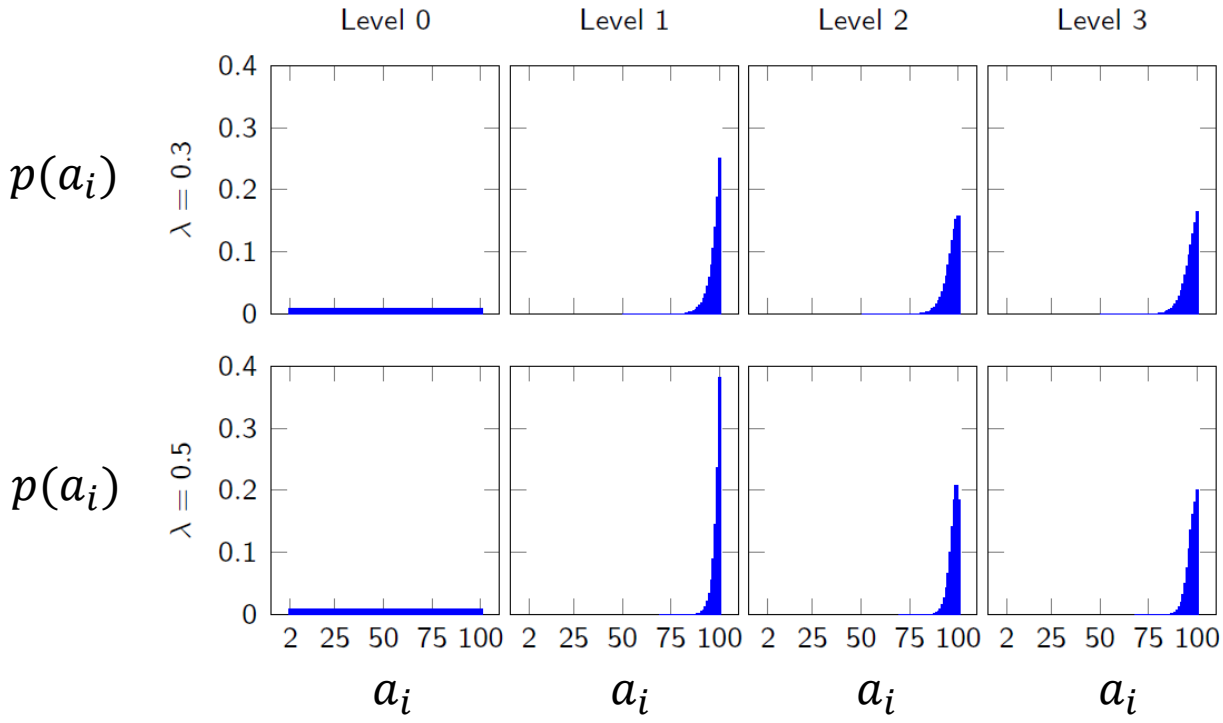
LUCY'S CHOICE (dollars) \ PETE'S CHOICE (dollars)	2	3	4	…	98	99	100
2	2 2	4 0	4 0	…	4 0	4 0	4 0
3	0 4	3 3	5 1	…	5 1	5 1	5 1
4	0 4	1 5	4 4	…	6 2	6 2	6 2
…	…	…	…		…	…	…
98	0 4	1 5	2 6	…	98 98	100 96	100 96
99	0 4	1 5	2 6	…	96 100	99 99	101 97
100	0 4	1 5	2 6	…	96 100	97 101	100 100

事实上，旅行者困境问题只有唯一一个纳什均衡：

$$(\$2, \$2)$$

旅行者困境的对率k层模型

- 有不同λ和k值的对率k层模型的策略



随着k值增加，两个不同λ值对应的策略之间的差异变得不明显

- 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：

$$VOI(O_j \mid \boldsymbol{o}) \geq 0 \quad \forall \boldsymbol{o}, O_j$$

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：

  $$P\big(pass(c_1,t_1)\mid q^+(c_1)\big)=0.8 \qquad P\big(pass(c_1,t_1)\mid q^-(c_1)\big)=0.35$$

  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的条件下，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，以及它们的期望效用，计算最优决策。

  （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

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：

2020年4月12日

本科生班的同学把作业发给范彧：
fanyu@lamda.nju.edu.cn

研究生班的同学把作业发给孔祥瀚：
kongxh@lamda.nju.edu.cn