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

2021年4月7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

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

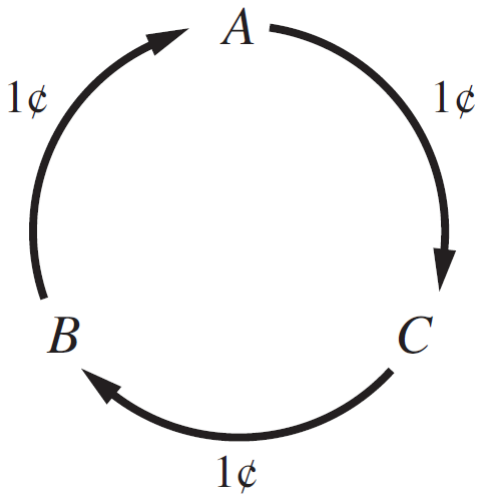
- **独立性：** 如果$A \succ B$，则对于任何C和概率p，有

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

- 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$P(s' \mid a, o)$：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**：

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

- **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

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

- 我们感兴趣的是理性Agent，MEU是智能系统设计的**中心原则**

例子

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

- 给定目的地的天气预报，应用最大期望效用原则决定是否带雨伞

概率模型$P(s' \mid a, o)$

o	a	s'	$P(s' \mid a, o)$
forecast rain	bring umbrella	rain with umbrella	0.9
forecast rain	leave umbrella	rain without umbrella	0.9
forecast rain	bring umbrella	sun with umbrella	0.1
forecast rain	leave umbrella	sun without umbrella	0.1
forecast sun	bring umbrella	rain with umbrella	0.2
forecast sun	leave umbrella	rain without umbrella	0.2
forecast sun	bring umbrella	sun with umbrella	0.8
forecast sun	leave umbrella	sun without umbrella	0.8

效用函数$U(s')$

s'	$U(s')$
rain with umbrella	−0.1
rain without umbrella	−1
sun with umbrella	0.9
sun without umbrella	1

- 如果天气预报为下雨，则带伞和不带伞的期望效用分别为

$$EU(\text{bring umbrella} \mid \text{forecast rain}) = 0.9 \times -0.1 + 0.1 \times 0.9 = 0$$

$$EU(\text{leave umbrella} \mid \text{forecast rain}) = 0.9 \times -1 + 0.1 \times 1 = -0.8$$

⇨ 应该带伞

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）

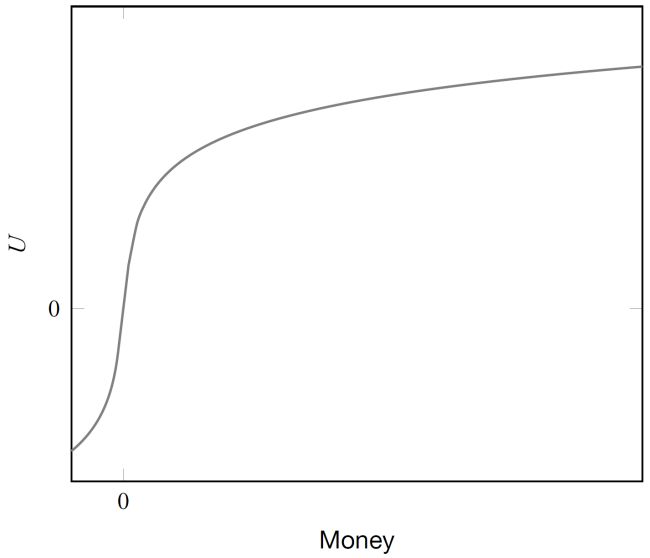
避碰示例

- 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

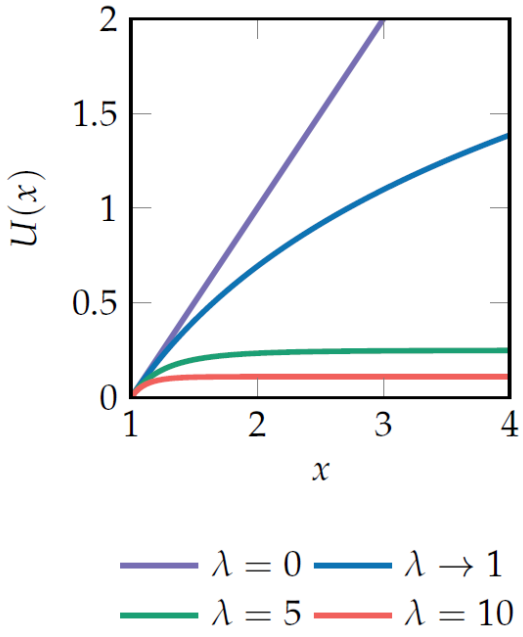
- 经验心理学的研究表明，$U(k) = \alpha + \beta \log(k + \gamma)$

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（$A \succ B$）

风险规避：幂效用



- 常用的风险规避函数：幂效用

$$U(x) = \frac{x^{1-\lambda} - 1}{1 - \lambda}$$

其中，$\lambda \geq 0$且$\lambda \neq 1$

- 对数效用$U(x) = \ln x$：幂效用在$\lambda \to 1$时的特殊形式

- 对于$x > 0$，$\lambda > 0$且$\lambda \neq 1$，幂效用函数$U(x)$是风险规避的：

$$\frac{d^2U}{dx^2} = \frac{-\lambda}{x^{\lambda+1}} < 0$$

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

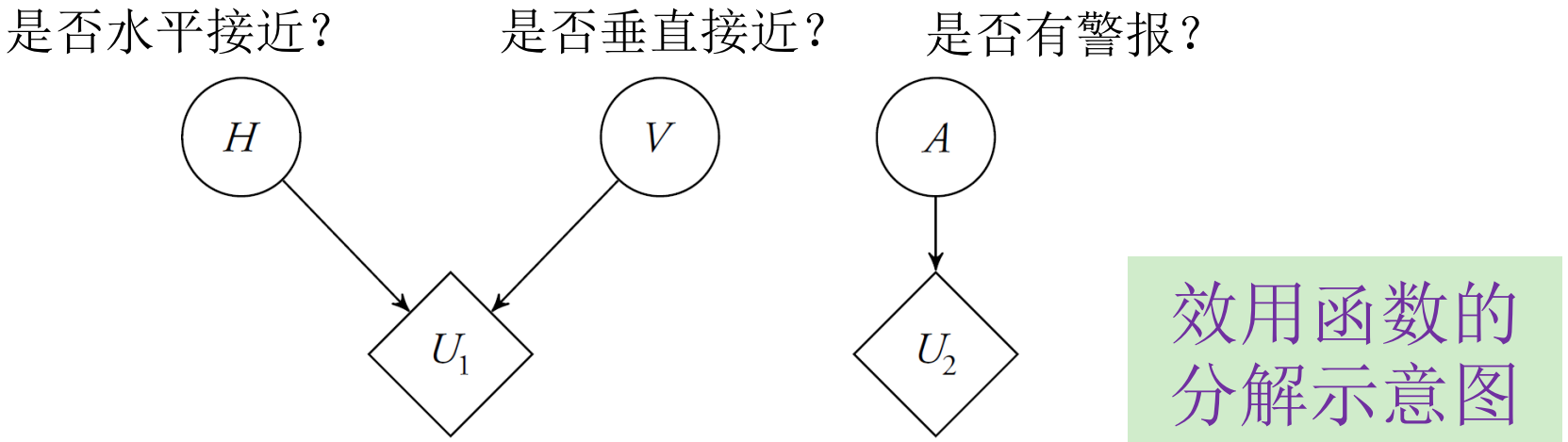
效用启发式

- 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）
- **机会结点**（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

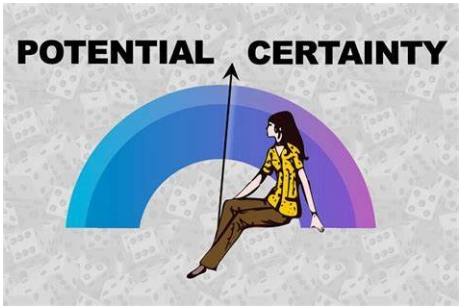
- 决策理论：一种规范性理论，描述了一个理性的Agent应该如何行动
- 描述性理论：描述了实际的Agent（例如人类）真正如何行动
- 有一些实验证据表明这两者不是一致的
- 人类的偏好在很多时候是非理性的

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、效用函数分解
- 最大化期望效用原则
- 人类评价与非理性
  - Allais悖论、Ellsberg悖论、表达效应、锚效应

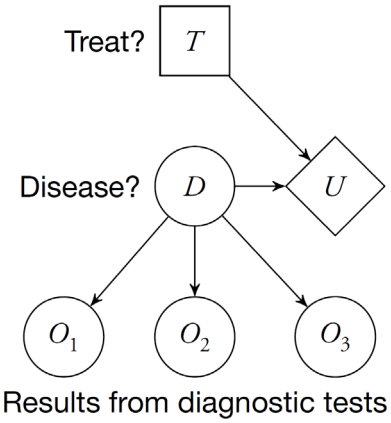
内容安排

- 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（圆形）：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

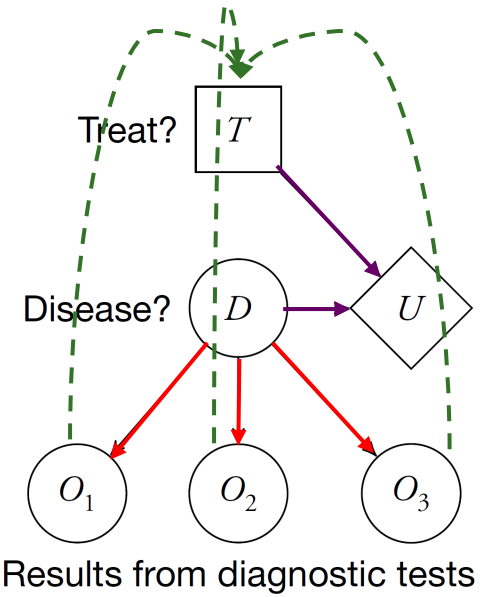
T	D	$U(T,D)$
0	0	0
0	1	−10
1	0	−1
1	1	−1

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

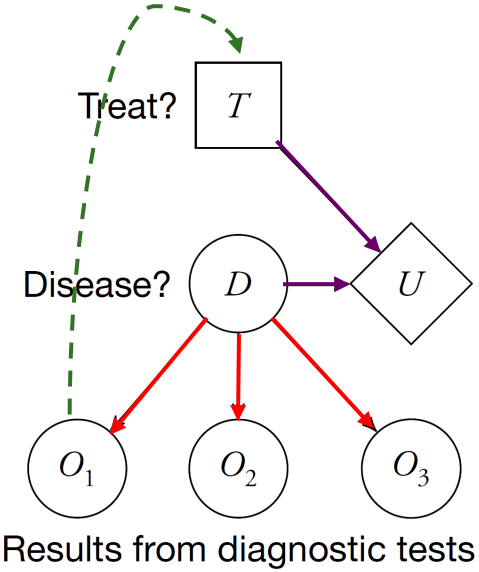
- **把决策问题表示为决策网络**
  - 利用问题的结构来计算基于效用函数的最优决策

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**

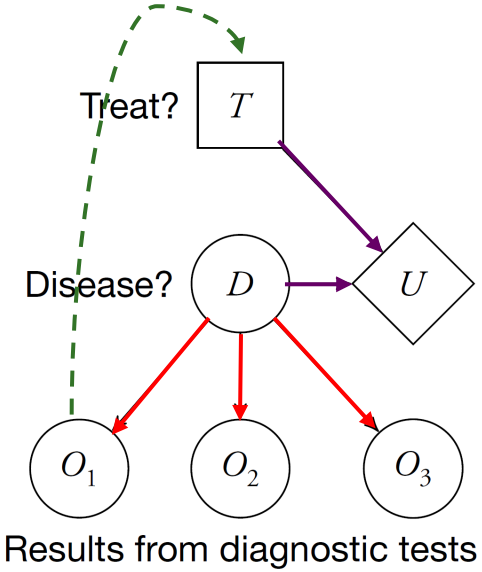
- 假设仅有第一次诊断测试的结果（正面的，记为o_1^1）
  - 添加一条从O_1到T的信息边，有

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

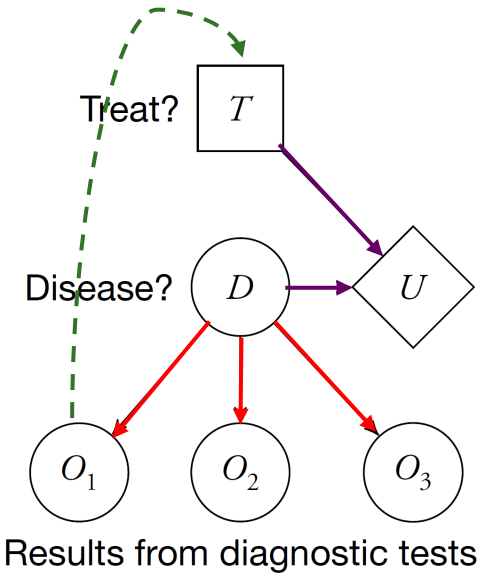
（2）对于决策结点的每个值：

（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

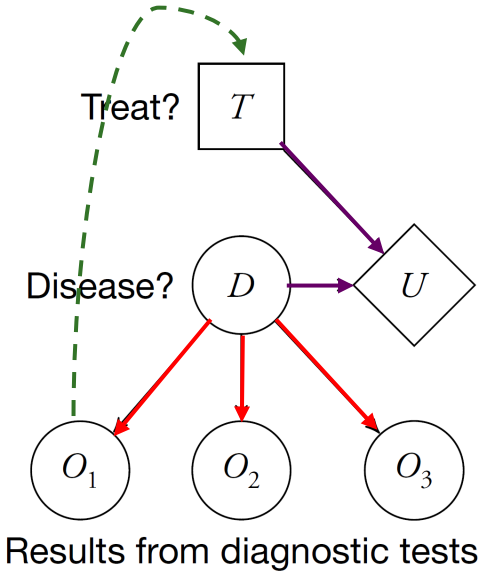
（3）返回最高效用的行动



- **改进算法：**如果决策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（圆形）
    - 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

- 假设$A \succcurlyeq C \succcurlyeq B$，每个结果的效用分别为$U(A) = 450$，$U(B) = -150$，$U(C) = 60$。试给出定义在$A$和$B$上的一次抽奖，使得它的效用与$C$的效用相同。

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？