

人工智能基础极限复习（救急版）

第一章

智能的定义（什么是智能、人工智能）

智能是在某种社会或文化环境的价值标准下，个体用以解决自己遇到的真正难题或生产及创造出有效产品所需要的能力。

人工智能的学派（特点，优缺点）

学派	特点	缺点
符号主义 强调物理符号系统 强调对知识的处理 强调 <u>知识表示</u>	1. 立足于逻辑运算和符号操作，适合于模拟人的逻辑思维过程,解决需要逻辑推理的复杂问题 2. 知识可用显示的符号表示,在已知基本规则的情况下，无需输入大量的细节知识 3. 便于模块化,当个别事实发生变化时，易于修改 4. 能与传统的符号数据库进行连接 5. 可对推理结论进行解释,便于对各种可能性进行选择	1. 有时体现的是“暴力”的思想（四色定理的证明） 2. 可以解决逻辑思维,但对于形象思维难于模拟
连接主义 又称仿生学派，强调 <u>神经元</u> 的运作，属于非符号处理范畴	1. 通过神经元之间的并行协作实现信息处理,处理过程具有并行性,动态性,全局性 2. 可以实现联想的功能,便于对有噪声的信息进行处理 3. 可以通过对神经元之间连接强度的调整实现学习和分类等 4. 适合模拟人类的形象思维过程 5. 求解问题时,可以较快的得到一个近似解	1. 不适合于解决逻辑思维 2. 黑盒 3. 费用高
行为主义 智能行为的基础是“感知—行动”，是在与环境的 <u>交互</u> 作用中表现出来的。比如强化学习	不需要知识，不需要表示，不需要推理	（课件里说尚未形成完整的理论体系，有待研究，引起注意，MD 什么时候的课件都过去多少年了早就形成一体系一大堆人在研究了）

第二章

谓词公式表示知识（看懂下面就行，怎么变为数学语言）（谓词用大写，函数变量用小写）

- 所有的人都是会死的， $\forall x(Human(x) \rightarrow Die(x))$
- 因为诸葛亮是人， $Human(Zhugeliang)$
- 所以诸葛亮是会死的。 $Die(Zhugeliang)$

$$(3) P, P \rightarrow Q \Rightarrow Q$$

- {1} $\forall x(Human(x) \rightarrow Die(x))$ **$P \rightarrow Q$**
- {2} $Human(Zhugeliang), Human(x) \rightarrow Die(x)$ **P 规则**
- {1, 2} $Human(Zhugeliang), Human(x) \rightarrow Die(x) \Rightarrow Die(Zhugeliang)$ **假言推理**

谓词公式表示知识的步骤：

- 定义谓词及个体。
- 变元赋值。
- 用连接词连接各个谓词，形成谓词公式。

产生式系统的构成（各部分作用）

1. 确定性规则知识的产生式表示

基本形式：IF P THEN Q

或者： $P \rightarrow Q$

例如：

r_4 : IF 动物会飞 AND 会下蛋 THEN 该动物是鸟

产生式：

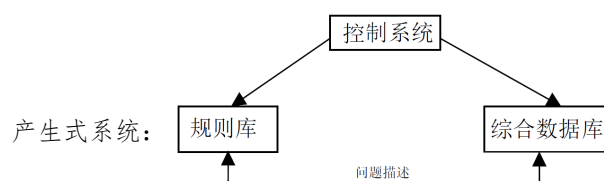
确定性的三元组：（对象，属性，值）

2. 不确定性规则知识的产生式表示

基本形式：IF P THEN Q （置信度）

或者： $P \rightarrow Q$ （置信度）

例如：IF 发烧 THEN 感冒（0.6）



产生式系统的基本结构

1. 规则库

规则库：用于描述相应领域内知识的产生式集合。

2. 综合数据库

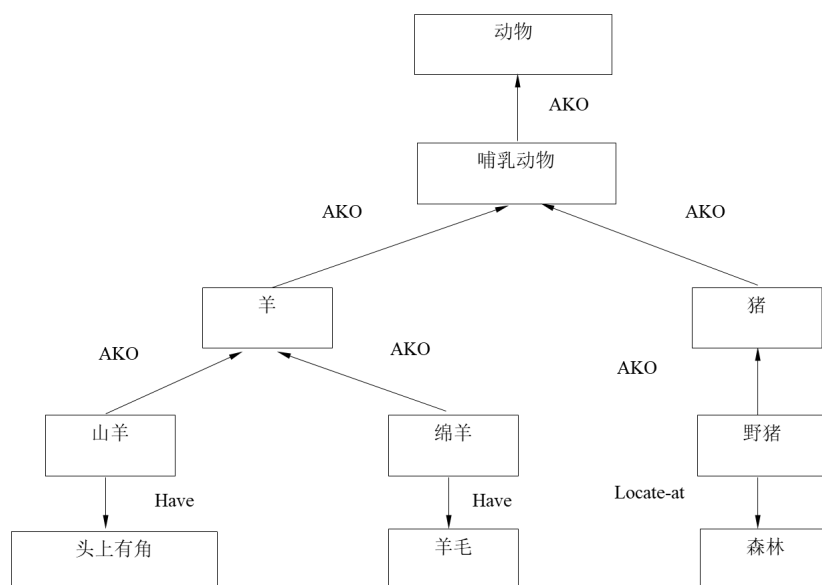
综合数据库（事实库、上下文、黑板等）：一个用于存放问题求解过程中各种当前信息的数据结构。

3. 控制系统

控制系统（推理机构）：由一组程序组成，负责整个产生式系统的运行，实现对问题的求解。

【必考】语义网络表示知识（连词量词的表示）

下面的例子看懂就行了大概长这样：



一些简写：

AKO: a kind of

AMO: a member of

ISA: is a

自己写写中文得了

第三章

★**归结演绎推理**（谓词演算等价式、子句和子句集、子句集化简、鲁宾逊归结原理、置换合一、归结反演）（化子句集或证明题）

子句：析取式

子句集：{子句 1, 子句 2}，逗号看成合取

双重否定律 $\neg(\neg P) \Leftrightarrow P$

德.摩根律 $\neg(P \wedge Q) \Leftrightarrow \neg P \vee \neg Q, \neg(P \vee Q) \Leftrightarrow \neg P \wedge \neg Q$

量词转换律 $\neg(\exists x)P \Leftrightarrow (\forall x)\neg P, \neg(\forall x)P \Leftrightarrow (\exists x)\neg P$

【谓词公式化为子句集】方法和例子

$(\forall x)\{[\neg P(x) \vee \neg Q(x)] \rightarrow (\exists y)[S(x, y) \wedge Q(x)]\} \wedge (\forall x)[P(x) \vee B(x)]$

- (1) 消去蕴含符号 $P \rightarrow Q \Leftrightarrow \neg P \vee Q$

$(\forall x)\{[\neg[\neg P(x) \vee \neg Q(x)] \vee (\exists y)[S(x, y) \wedge Q(x)]] \wedge (\forall x)[P(x) \vee B(x)]$

- (2) 把否定符号移到每个谓词前面

$(\forall x)\{[P(x) \wedge Q(x)] \vee (\exists y)[S(x, y) \wedge Q(x)]\} \wedge (\forall x)[P(x) \vee B(x)]$

- (3) 变量标准化

$(\forall x)\{[P(x) \wedge Q(x)] \vee (\exists y)[S(x, y) \wedge Q(x)]\} \wedge (\forall w)[P(w) \vee B(w)]$

- (4) 消去存在量词，设 y 的函数是 $f(x)$ ，则

$(\forall x)\{[P(x) \wedge Q(x)] \vee [S(x, f(x)) \wedge Q(x)]\} \wedge (\forall w)[P(w) \vee B(w)]$

- (5) 化为前束形

$(\forall x)(\forall w)\{[P(x) \wedge Q(x)] \vee [S(x, f(x)) \wedge Q(x)]\} \wedge [P(w) \vee B(w)]$

- (6) 化为标准形

$P \vee (Q \wedge R) \Leftrightarrow (P \vee Q) \wedge (P \vee R)$
 $P \wedge (Q \vee R) \Leftrightarrow (P \wedge Q) \vee (P \wedge R)$

$(\forall x)(\forall w)\{[Q(x) \wedge P(x)] \vee [Q(x) \wedge S(x, f(x))]\} \wedge [P(w) \vee B(w)]$

$(\forall x)(\forall w)\{Q(x) \wedge [P(x) \vee S(x, f(x))]\} \wedge [P(w) \vee B(w)]$

- (7) 略去全称量词

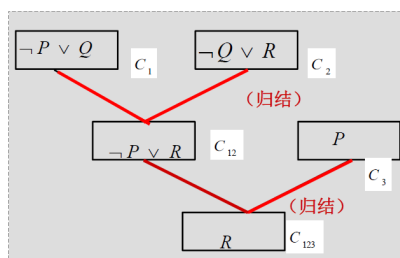
$Q(x) \wedge [P(x) \vee S(x, f(x))] \wedge [P(w) \vee B(w)]$

- (8) 消去合取词，把母式用子句集表示

$\{Q(x), P(x) \vee S(x, f(x)), P(w) \vee B(w)\}$

- (9) 子句变量标准化 $\{Q(x), P(y) \vee S(y, f(y)), P(w) \vee B(w)\}$

【鲁滨逊归结原理】(你可能会推出悖论出来，比如 PQR 是 100 的时候结果是假.我们不管，子句集中的句子一定是真的。比如{x 是电器，x 在室内，x 制冷效果好}的子句集，研究范围是电器、室内，我们为什么要去照顾“x 是洗衣机”“x 在室外”的情况，也就是说真值表中结果为 0 的部分是 100%不会发生的，所以随便给它赋值都无所谓，就是说 1 一定要是 1 但是 0 可以变成 1。其实应付考试只需要把 $\neg P \vee Q$ 看成 $P \rightarrow Q$ 然后当作数学题做就可以了)



【置换合一】如果研究对象是美的空调，直接取 x =美的空调，替换掉 x ；有的时候可能要 $x=g(t)$ 之类的，如果 x 和 t 有关的话，比如 t 表示厂家所在的国家，研究国产空调，就要取 t =中国， $x=g(t)$ 。

【归结反演】

你有条件 A, B, C, 要推出结论 Q, 那证明子句集 $\{A, B, C, \neg Q\}$ 的归结结果是空的就行。

证明: $\{A, B, C\}$ 是 $\{Q\}$, 那 $\{Q, \neg Q\}$ 不就是空吗

举例子就会了:

★ 例3.9 某公司招聘工作人员, A, B, C 三人应试, 经面试后公司表示如下想法:

- (1) 三人中至少录取一人。
- (2) 如果录取 A 而不录取 B, 则一定录取 C。
- (3) 如果录取 B, 则一定录取 C。

- 求证: 公司一定录取 C。

★ 证明：公司的想法用谓词公式表示： $P(x)$ ：录取 x 。

- (1) $P(A) \vee P(B) \vee P(C)$
- (2) $P(A) \wedge \neg P(B) \rightarrow P(C)$
- (3) $P(B) \rightarrow P(C)$

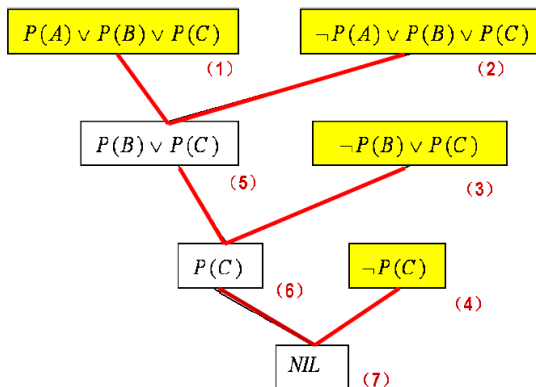
■ 把要求证的结论用谓词公式表示出来并否定，得：

- (4) $\neg P(C)$

$$P \rightarrow Q \Leftrightarrow \neg P \vee Q$$

■ 把上述公式化成子句集：

- (1) $P(A) \vee P(B) \vee P(C)$
- (2) $\neg P(A) \vee P(B) \vee P(C)$
- (3) $\neg P(B) \vee P(C)$
- (4) $\neg P(C)$



(是不是很熟悉？拒取式，厉害吧，概率论里面也见到过)

第四章

状态空间 (知道概念)：二元组：state, operation

盲目搜索 (宽度、深度) (open 表、close 表动态变化情况)

(2) 一般图搜索算法★

□ 符号说明：

- **s**-初始状态节点
- **G**-搜索图
- **OPEN**-存放待扩展节点的表
- **CLOSE**-存放已被扩展的节点的表
- **MOVE-FIRST(OPEN)**-取OPEN表首的节点作为当前要被扩展的节点n，同时将节点n移至CLOSE表

□ 一般图搜索算法划分为二个阶段：

- 1、初始化
- 2、搜索循环

其实就是 BFS 和 DFS，秒了

★ 启发式搜索 (评估函数、A 与 A* 算法) (带星号就代表就是最优的解，最短路径，不是估计)

看例题和作业题的八字码游戏，反正就是深搜但是优先搜索评估函数值最小的

A 算法评价函数： $f(n) = g(n) + h(n)$ ， g 是已经走过的实际的路径代价， h 是还没走过的估计的最小路径代价

A 算法的关键是 $h(n)$ 尽可能靠近 $h^*(n)$

A* 算法是可采纳的，总能搜索到最短解答路径。

A* 算法搜索问题解答的关键： $h(n)$ 在满足 $h(n) \leq h^*(n)$ 的条件下，越大越好

问题规约 (AND-OR 图、解图、极大极小过程、 α - β 过程)

问题规约三元组： (S_0, O, P) ，subject / operator / primitive problem set，主问题、操作算子、公理集

【与或图】同一层级，括弧括起来的是“与”，它们是小团体，一定要在一起，不同小团体之间是“或”

【解图】没有或的 AND-OR 图，比如右边那个

与或图的启发式搜索 AO* (搜索解图)：b23.tv/av798137588 (解题从 2:10)

(博弈树中，自己的局是或，对方的局是与)

极大极小过程： $f(p)$ ，利正劣负，自己的局取极大，对方的局取极小

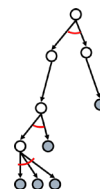
α - β 剪枝：解决博弈树组合爆炸问题，所以要边生成博弈树，边进行估值的计算 b23.tv/av286202006



与或图 AO* 算法



α - β 剪枝过程



第五章（計算問題があります，絶対に）

主观贝叶斯方法（区分几率和概率、几率表示证据、组合证据的计算、不确定性的更新，证据为真/假对应的公式）

【Bayse 公式】 $p(Q/P) = \frac{p(P/Q) \cdot p(Q)}{p(P)}$ ，我喜欢这么记： $p(Q/P) \cdot p(P) = p(P/Q) \cdot p(Q)$

$$\begin{aligned} p(Q/P) \cdot p(P) &= p(P/Q) \cdot p(Q) \\ p(\neg Q/P) \cdot p(P) &= p(P/\neg Q) \cdot p(\neg Q) \end{aligned} \xrightarrow{Q \leftrightarrow \neg Q} \begin{cases} \frac{p(Q/P)}{p(\neg Q/P)} = \frac{p(P/Q)}{p(P/\neg Q)} \cdot \frac{p(Q)}{p(\neg Q)} \\ O(Q/P) = LS \cdot O(Q) \end{cases} \xrightarrow{P \leftrightarrow \neg P} \begin{cases} \frac{p(Q/\neg P)}{p(\neg Q/\neg P)} = \frac{p(\neg P/Q)}{p(\neg P/\neg Q)} \cdot \frac{p(Q)}{p(\neg Q)} \\ O(Q/\neg P) = LN \cdot O(Q) \end{cases}$$

LS——充分性因子

- $=1: O(Q/P)=O(Q)$, P对Q无影响;
- $>1: O(Q/P)>O(Q)$, P支持Q;
- $<1: O(Q/P)<O(Q)$, P不支持Q;

LN——必要性因子

- $=1: O(Q/\neg P)=O(Q)$, $\neg P$ 对Q无影响;
- $>1: O(Q/\neg P)>O(Q)$, $\neg P$ 支持Q;
- $<1: O(Q/\neg P)<O(Q)$, $\neg P$ 不支持Q;

可信度方法（不确定性推理算法、综合可信度计算）

1. 知识不确定性的表示

产生式规则表示:

IF E THEN H ($CF(H,E)$)

$CF(H,E)$: 可信度因子 (certainty factor), 反映前提条件与结论的联系强度。

$$CF(H) = CF(H,E) \cdot \max\{0, CF(E)\}$$

IF 头痛 AND 流涕 THEN 感冒 (0.7)

3. 组合证据不确定性的算法

组合证据: 多个单一证据的合取

$$E=E_1 \text{ AND } E_2 \text{ AND } \cdots \text{ AND } E_n$$

则 $CF(E)=\min\{CF(E_1), CF(E_2), \dots, CF(E_n)\}$

组合证据: 多个单一证据的析取

$$E=E_1 \text{ OR } E_2 \text{ OR } \cdots \text{ OR } E_n$$

则 $CF(E)=\max\{CF(E_1), CF(E_2), \dots, CF(E_n)\}$

5. 结论不确定性的合成算法

由多条不同的知识推出了相同的结论，求综合可信度。

设知识: IF E_1 THEN H ($CF(H,E_1)$)
IF E_2 THEN H ($CF(H,E_2)$)

(1) 分别对每一条知识求出 $CF(H)$:

$$CF_1(H) = CF(H, E_1) \times \max\{0, CF(E_1)\}$$

$$CF_2(H) = CF(H, E_2) \times \max\{0, CF(E_2)\}$$

(2) 求出 E_1 与 E_2 对 H 的综合影响所形成的可信度 $CF_{1,2}(H)$:

$$CF_{1,2}(H) = \begin{cases} CF_1(H) + CF_2(H) - CF_1(H)CF_2(H) & \text{若 } CF_1(H) \geq 0, CF_2(H) \geq 0 \\ CF_1(H) + CF_2(H) + CF_1(H)CF_2(H) & \text{若 } CF_1(H) < 0, CF_2(H) < 0 \\ \frac{CF_1(H) + CF_2(H)}{1 - \min\{|CF_1(H)|, |CF_2(H)|\}} & \text{若 } CF_1(H) \text{ 与 } CF_2(H) \text{ 异号} \end{cases}$$

证据理论（概率分配函数、信任函数、似然函数、类概率函数（核心）的概念与公式）

设 $D=\{\text{红}, \text{黄}, \text{蓝}\}$

$M(\{\text{红}\})=0.3, M(\{\text{黄}\})=0, M(\{\text{蓝}\})=0.1,$

概率分配函数: $M(\{\text{红}, \text{黄}\})=0.2, M(\{\text{红}, \text{蓝}\})=0.2,$ (和概率不同)

$M(\{\text{黄}, \text{蓝}\})=0.1, M(\{\text{红}, \text{黄}, \text{蓝}\})=0.1, M(\Phi)=0$

但: $M(\{\text{红}\}) + M(\{\text{黄}\}) + M(\{\text{蓝}\}) = 0.4 \neq 1$

信任函数:

$$Bel(D) = \sum_{B \subseteq D} M(B) = 1$$

设 $D=\{\text{红}, \text{黄}, \text{蓝}\}$

$M(\{\text{红}\})=0.3, M(\{\text{黄}\})=0, M(\{\text{红}, \text{黄}\})=0.2,$

$Bel(\{\text{红}, \text{黄}\})=M(\{\text{红}\})+M(\{\text{黄}\})+M(\{\text{红}, \text{黄}\})=0.3+0.2=0.5$

似然函数:

$$Pl(A) = 1 - Bel(\neg A)$$

设 $D=\{\text{红}, \text{黄}, \text{蓝}\}$

$M(\{\text{红}\})=0.3, M(\{\text{黄}\})=0, M(\{\text{红}, \text{黄}\})=0.2,$

$$Bel(\{\text{红}, \text{黄}\}) = M(\{\text{红}\}) + M(\{\text{黄}\}) + M(\{\text{红}, \text{黄}\}) = 0.3 + 0.2 = 0.5$$

$$Pl(\{\text{蓝}\}) = 1 - Bel(\neg\{\text{蓝}\}) = 1 - Bel(\{\text{红}, \text{黄}\}) = 1 - 0.5 = 0.5$$

定义4.4 设 M_1 和 M_2 是两个概率分配函数；则其正交和 $M = M_1 \oplus M_2$ ： $M(\Phi) = 0$

$$M(A) = K^{-1} \sum_{x \cap y = A} M_1(x) M_2(y)$$

概率分配函数正交：

$$\text{其中：} K = 1 - \sum_{x \cap y = \Phi} M_1(x) M_2(y) = \sum_{x \cap y \neq \Phi} M_1(x) M_2(y)$$

如果 $K \neq 0$ ，则正交和 M 也是一个概率分配函数；

如果 $K = 0$ ，则不存在正交和 M ，即没有可能存在概率函数，称 M_1 与 M_2 矛盾。

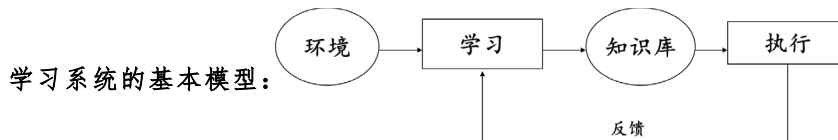
类概率函数：
$$f(A) = Bel(A) + \frac{|A|}{|\Omega|} [Pl(A) - Bel(A)]$$

$$= \frac{|A|}{|\Omega|} Pl(A) + \frac{|\Omega| - |A|}{|\Omega|} Bel(A)$$
，用来衡量证据 A 的不确定性的

第六章

机器学习概念（学习与机器学习，学习系统的基本要求、基本模型、四部分构成(各部分简单概念)）

机器学习：让机器来模拟人类的学习功能，是一门研究怎样用机器来模拟或实现人类学习活动的一门学科



环境向系统的学习部分提供某些信息。（影响学习系统设计的最重要因素）

学习部分利用这些信息修改知识库，以增进系统执行部分完成任务的效能，

执行部分根据知识库完成任务，同时把获得的信息反馈给学习部分。

在具体的应用中，环境、知识库和执行部分决定了具体的工作内容，学习部分所需要解决的问题完全由上述三部分确定。

机器学习的算法分类（基于……的分类）

按照使用函数的不同	线性模型、非线性模型
按照学习准则的不同	统计方法和非统计方法
按照训练样本提供的信息以及反馈方式的不同	监督学习、无监督学习、强化学习

归纳学习（了解）

比如，男娘和女人聚在一起，让机器去区分男女。学习时先打好标签，然后把他们的头发长度、脸型、声音、性征什么的属性都拿出来。

□ 在专家系统中经常遇到

□ 示例学习任务：

■ 从一系列示例出发：

□ 正例；

□ 反例；

■ 生成一个反映这些示例本质的定义（概念描述）：

□ 覆盖所有的正例，而不包含任何反例；

□ 可用来指导对新例子的分类识别；

解描述



例子空间：男娘+女人

假设空间（长发）：头发长的人

假设空间（短发）：头发短的人

①特化搜索 最大的假设空间，不覆盖这个类别（找反例）

- 从最泛化的假设（概念描述）出发；
- 每次取用一个新的例子，产生一些特化的描述；
- 直到产生出足够特化的解描述；

②泛化搜索 最小的假设空间，能覆盖这个类别（找正例）

- 从最特化的假设（例子空间中的一个正例）开始；
- 每次取用一个新的例子，产生一些泛化的描述；
- 直到产生出足够泛化的解描述。

决策树学习 (ID3 算法)

信息量：概率大，信息量小；概率小，信息量大。两个独立事件的联合信息量等于各之和。

【信息量】 $I = -\log(p)$

【信源的信息量：信息熵】 $H(X) = \sum_i p(x_i) I(x_i) = -\sum_i p(x_i) \log p(x_i)$ ，每个符号信息量的统计平均值

ID3 算法：使分类时平均的测试次数（测试传递最小的信息量）最小



决策树 ID3 算法

ID3 算法的过程：b23.tv/av114754714606455

1. 分类能力最好的属性被选作树的根节点
2. 根节点的每个可能值产生一个分支
3. 训练样例被排列到适当的分支
4. 重复上面的过程

【信息增益】作为分类能力：信息增益小，分类能力好

对于一个属性 i ， $\text{Gain}(S, i) = E(S) - E(i)$ ($E(i)$ 以属性 i 为符号的信息熵， S 为目标类别)

“高”的分支的所需期望信息量为： $M_{\text{高度(高)}}$
 $-(2/5) \log_2 (2/5) - (3/5) \log_2 (3/5) = 0.971 \text{ bits}$

“矮”的分支的所需期望信息量为： $M_{\text{高度(矮)}}$
 $-(1/3) \log_2 (1/3) - (2/3) \log_2 (2/3) = 0.918 \text{ bits}$

C 以属性“高度”作划分后进一步判别所需的期望信息量为：
 $E_{\text{高度}} = 5/8 \times M_{\text{高度(高)}} + 3/8 \times M_{\text{高度(矮)}} = 0.951 \text{ bits}$

$$I = E(S) = -\frac{3}{8} \log_2 \frac{3}{8} - \frac{5}{8} \log_2 \frac{5}{8} = 0.954$$

$$\text{Gain}(S, \text{高度}) = E(S) - E(\text{高度}) = 0.003$$

高度	发色	眼睛	类别
矮	黑色	蓝色	-
高	黑色	蓝色	-
矮	金色	蓝色	+
高	金色	棕色	-
高	黑色	棕色	-
矮	金色	棕色	-
高	金色	蓝色	+
高	红色	蓝色	+

高	矮
{高, 金, 棕: -} {高, 红, 蓝: +} {高, 黑, 蓝: -} {高, 金, 蓝: +} {高, 黑, 棕: -}	{矮, 金, 蓝: +} {矮, 黑, 棕: -}