

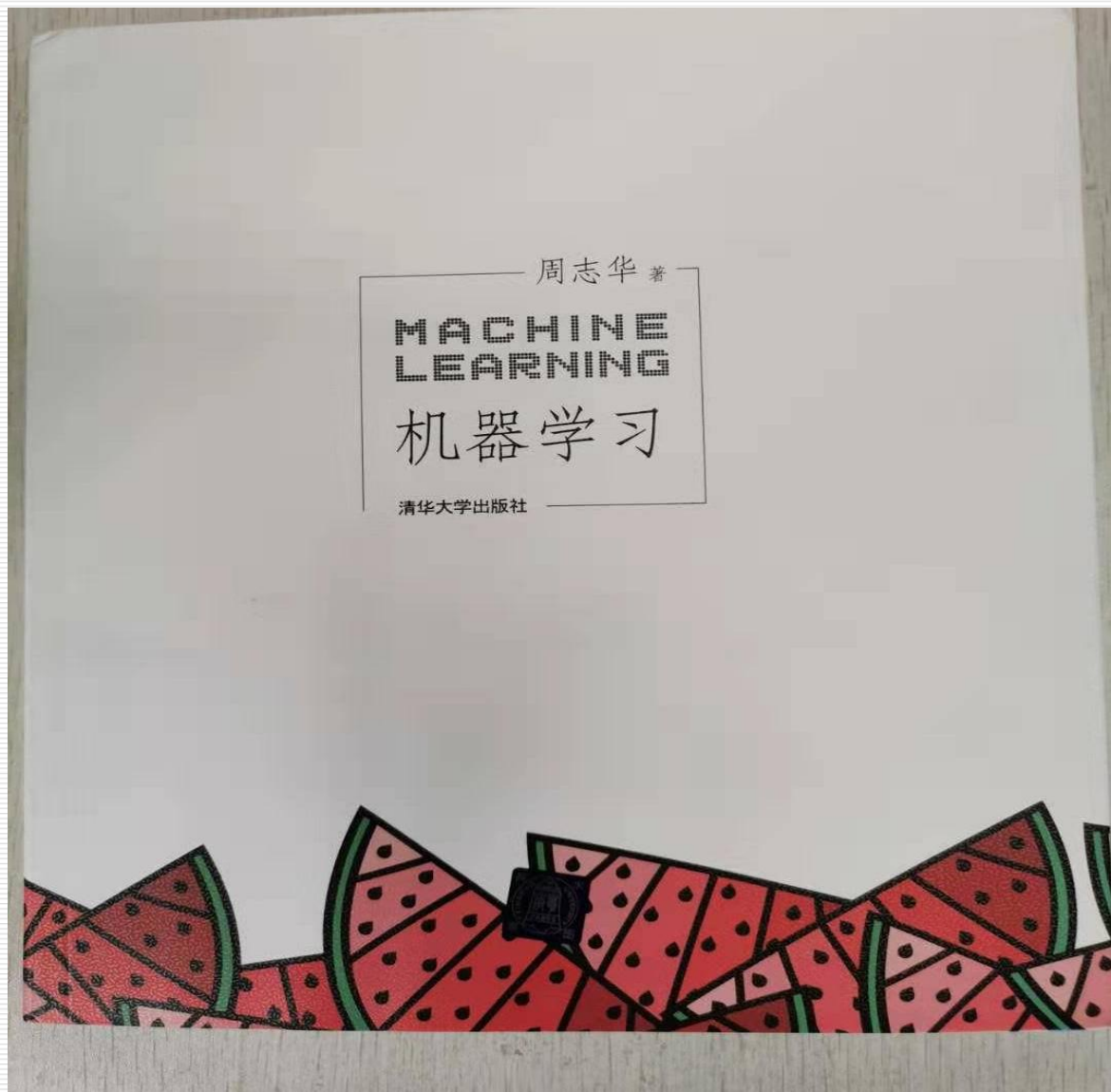
机器学习

陈志华

主要内容

- 机器学习的定义
- 机器学习的模型
- 决策树学习
- 概念学习

推荐书籍



前言

- 机器学习(Machine Learning)是一门多领域交叉学科，涉及概率论、统计学、逼近论、凸分析、算法复杂度理论等多门学科
- 机器学习研究计算机怎样模拟或实现人类的学习行为，以获取新的知识或技能，重新组织已有的知识结构使之不断改善自身的性能

前言

□ 机器学习算法

- 主要是设计和分析让计算机可以自动“学习”的算法
- 一类从数据中自动分析获得规律，并利用规律对未知数据进行预测的算法
- 由于涉及了大量的统计学理论，也被称为统计学习理论

前言

- 机器学习是一个过程，允许机器通过指令的接收或经验的积累对自身性能进行改进
- 动物学习具有不同的层次
 - 条件反射：刺激→反应
 - 人类学习：创新

前言

□ 学习系统的共同特征

- 具有一定的结构
- 需要训练
- 效果可以验证：通过解决训练问题之外的问题

机器学习的发展阶段

- 通用的学习系统研究（**50年代中期**）
 - 致力于构造一个没有或者只有很少初始知识的通用系统
 - 主要技术有神经元模型、决策论和控制论。
- 概念学习系统研究（**60年代中期**）
- 基于知识的各种学习系统研究（**70年代中期**）
- 联接学习和符号学习的深入研究（**80年代末**）
- 统计学习（**90年代中**）
- 深度学习（**00年代中-现在**）

学习的分类

□ 基于系统性的分类

- 归纳学习

- 分析学习

- 联结学习

- 遗传学习

机器学习研究内容

□ 通用学习算法

- 理论分析

- 开发用于非实用的学习任务算法。

□ 认知模型

- 研究人的学习的计算理论和实验模型。

□ 工程目标

- 解决专门的实际问题

- 开发完成这些任务的工程系统

机器学习研究内容

- 机器学习系统是在一定程度上实现机器学习的软件
 - 如果一个系统能够从某个过程或环境的未知特征中学习到有关信息，并能把学到的信息用于对未来的估计、分类、决策或控制，以便改进系统的性能，那么它就是一个学习系统

1. 定义

- 如果一个计算机程序针对某类任务T，用P衡量性能，根据经验E来自我完善，那么就称该程序在从经验E中学习，针对任务T，它的性能用P来衡量(Mitchell 97)
- 学习包括三个部分：
 - 任务种类T，性能标准P，经验来源E

例子

□ 例1：象棋学习问题

- 任务T: 和人类下象棋

- 性能标准P: 战绩

- 训练经验E: 和人对弈、自己对弈或打谱

例子

□ 例2：汉字手写体识别学习问题

- 任务T：识别和分类图像中的手写文字
- 性能标准P：识别和分类的正确率
- 训练经验E：已识别的手写体汉字的数据库

例子

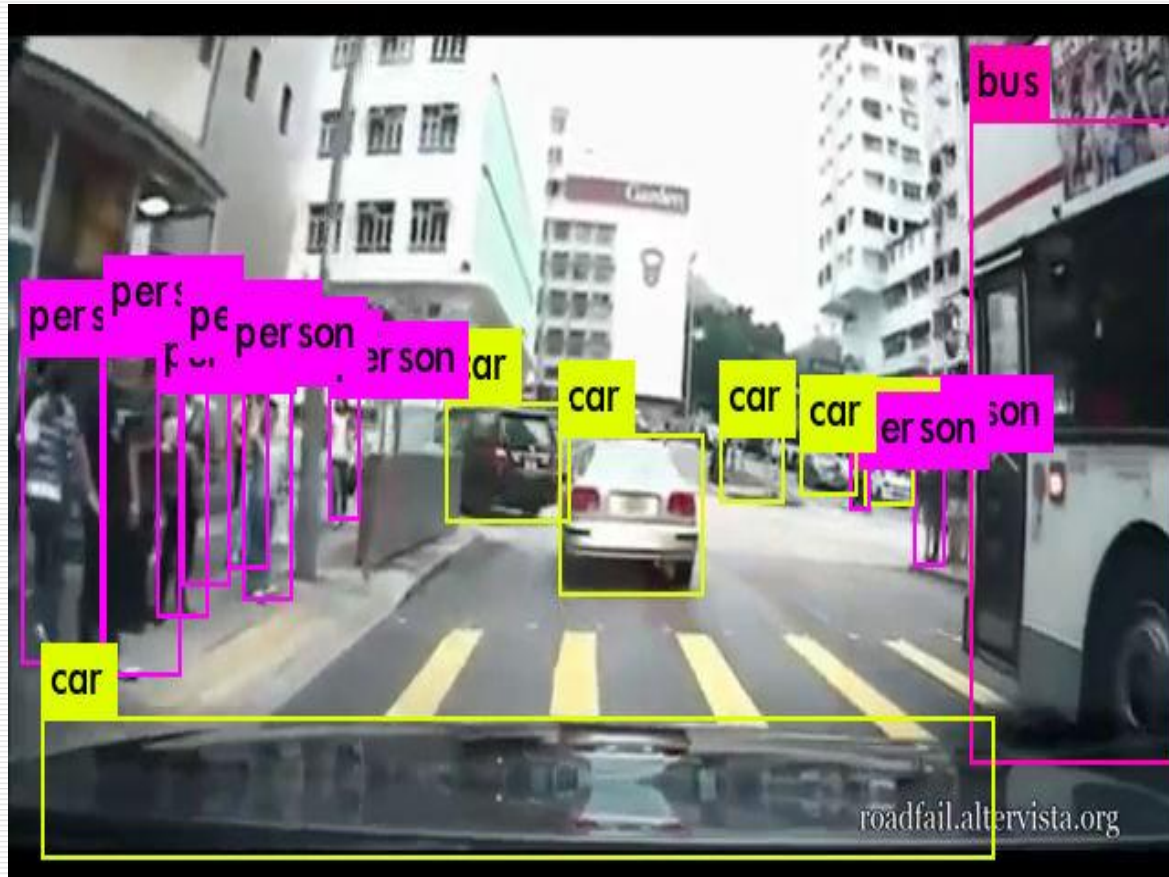
□ 例3：自动驾驶的学习问题

- T: 通过视觉传感器等在四车道高速公路上驾驶
- P: 平均无差错行驶里程
- E: 人类驾驶时录制的一系列图像和驾驶指令数据库

实例：自动驾驶—行人与移动目标检测

使用YOLO v3模型

对行人目标以及常见移动目标进行实时检测



实例：自动驾驶——车道线检测

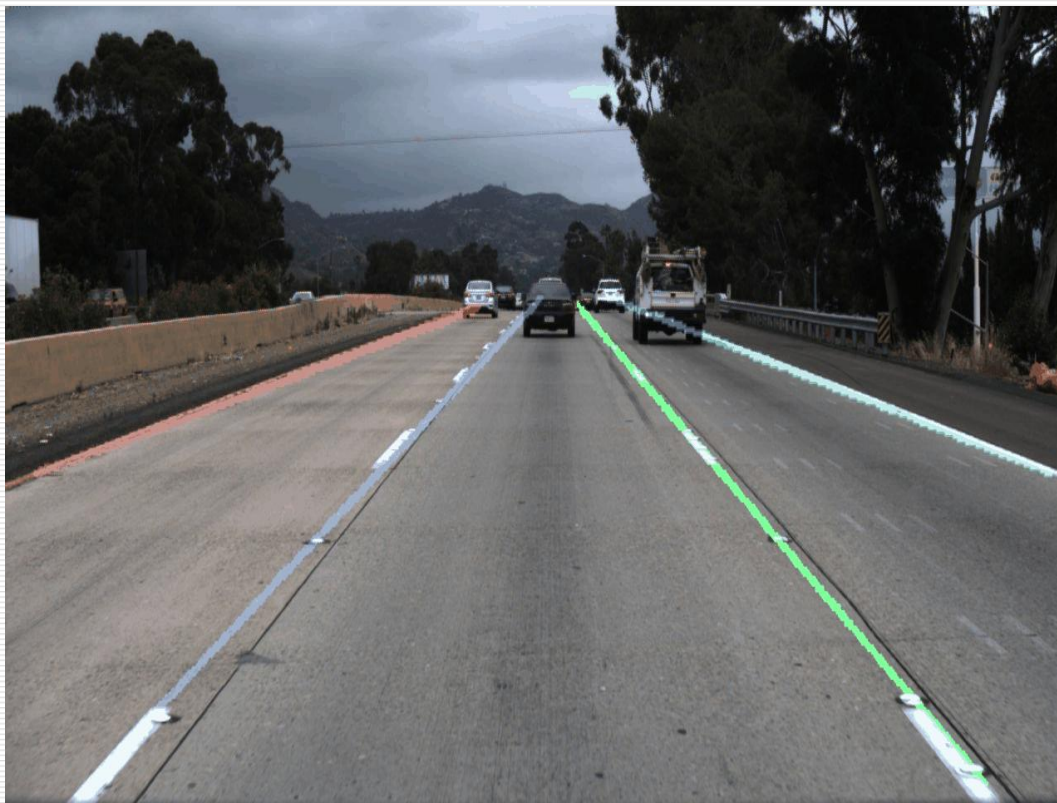


实例：自动驾驶——车位线检测

□ RCNN-Vehicle-Tracking-Lane-Detection



实例：自动驾驶——车位线检测



□ 学习系统的质量取决于

- 学习系统的结构

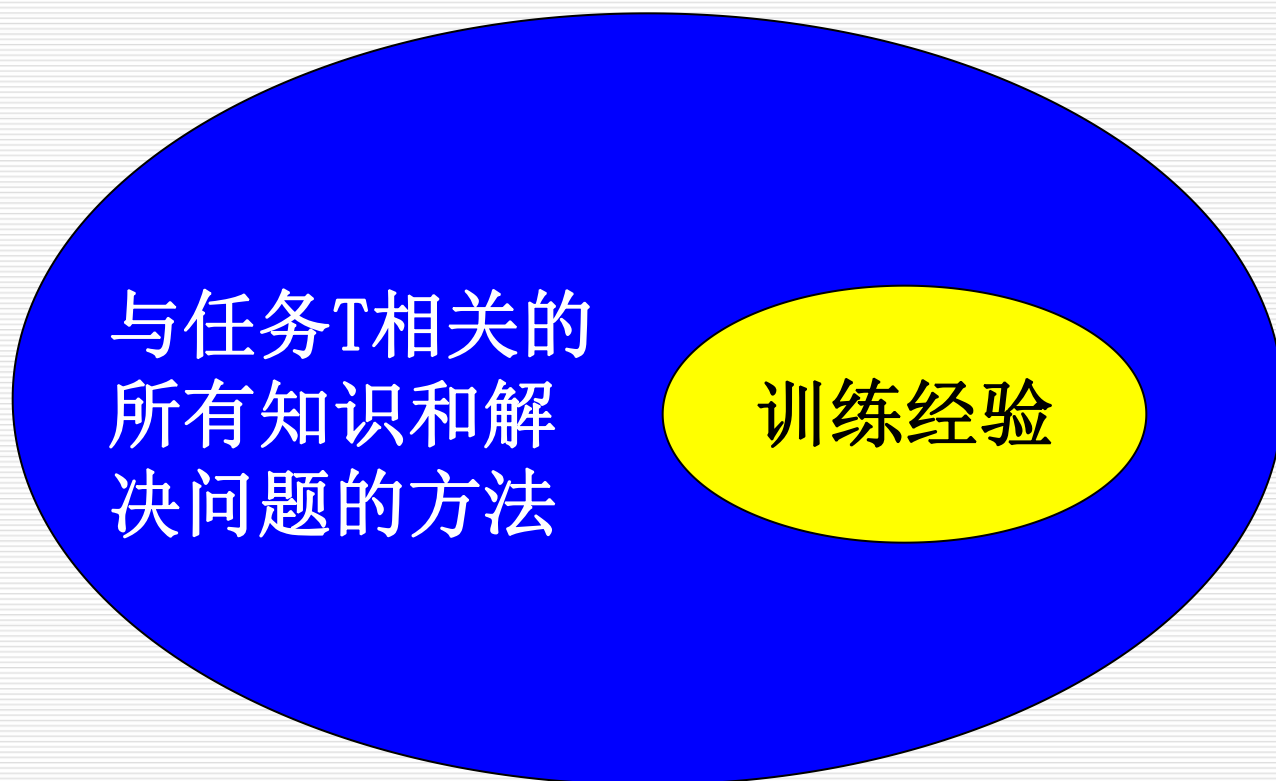
- 训练经验 E

□ 对应于机器学习系统

- 算法

- 样例

2. 机器学习模型



2. 机器学习模型

□ 学习模型的形式化表示

- (1) 实例集合 X
- (2) 假设集合 H
- (3) 训练样例集合 D
- (4) 目标函数 $C(x)$
- (5) 学习过程
- (6) 机器学习假设

(1) 实例集合X

- 与任务T相关的所有知识和解决问题的方法
- 可以包括正例(正确的)和反例（错误的）
- 例：任务T是“100以内的整数乘法”，则实例集合就是0到100之间的整数之间进行相乘的所有情形
 - 正例： $10 \times 20 = 200$ 等等；
 - 反例： $10 \times 20 = 100$ 等等

(2) 假设集合H

□ 学习系统的学习结果的集合

□ 表示方法包括：

- 属性表

- 规则

- 决策树

(3) 训练样例集合D

- 从实例集合中选出，用于训练学习系统进行学习的
样例集合
- 样例质量对学习系统的学习效果影响很大
 - 如教材

(4) 目标函数 $C(x)$

□ 是一个理想函数，参数 x 来自实例集合

□ 取值：

$$C(x) = \begin{cases} 1, & \text{当 } x \text{ 是正例} \\ 0, & \text{当 } x \text{ 是反例} \end{cases}$$

(5) 学习过程

- 是一个在假设集合 H 中进行搜索的过程
- 学习过程使得搜索到的假设在训练样例集上与 $C(x)$ 一致：若 $C(x)$ 将某训练样例判定为正（反）例，则学习系统学习到的假设也必须判为正（反）例

(5) 学习过程

□ 学习过程的表示

对于 $x \in D$ (训练样例集合), 有: $h(x) = C(x)$

□ 未知的与任务T相关的知识和解决方案可以表示为: 实例集合 **X**-训练样例集合 **D**

(5) 学习过程

- 学习系统学习到的知识，就是在 D 上与 $C(x)$ 保持一致的所有假设
- 检查学习效果，就是应用在未知集合 $X-D$ 时，检查学习系统是否能得到正确结果，也就是在 $X-D$ 上与 $C(x)$ 是否保持一致
- 性能 P :

学习系统在 $X-D$ 上与 $C(x)$ 保持一致的实例数目

(6) 机器学习的假设规律

- 学习系统通过学习获得的假设，如果在 D 上逼近目标函数 $C(x)$ ，那么也能在 $X-D$ 上逼近目标函数 $C(x)$

3. 决策树学习

- 决策树学习是人们广泛使用的一种归纳推理形式，
它需要给定一组训练样例，其中每个训练样例都已
标记上正例和反例
- 根据这组训练样例，可以创建相应的决策树

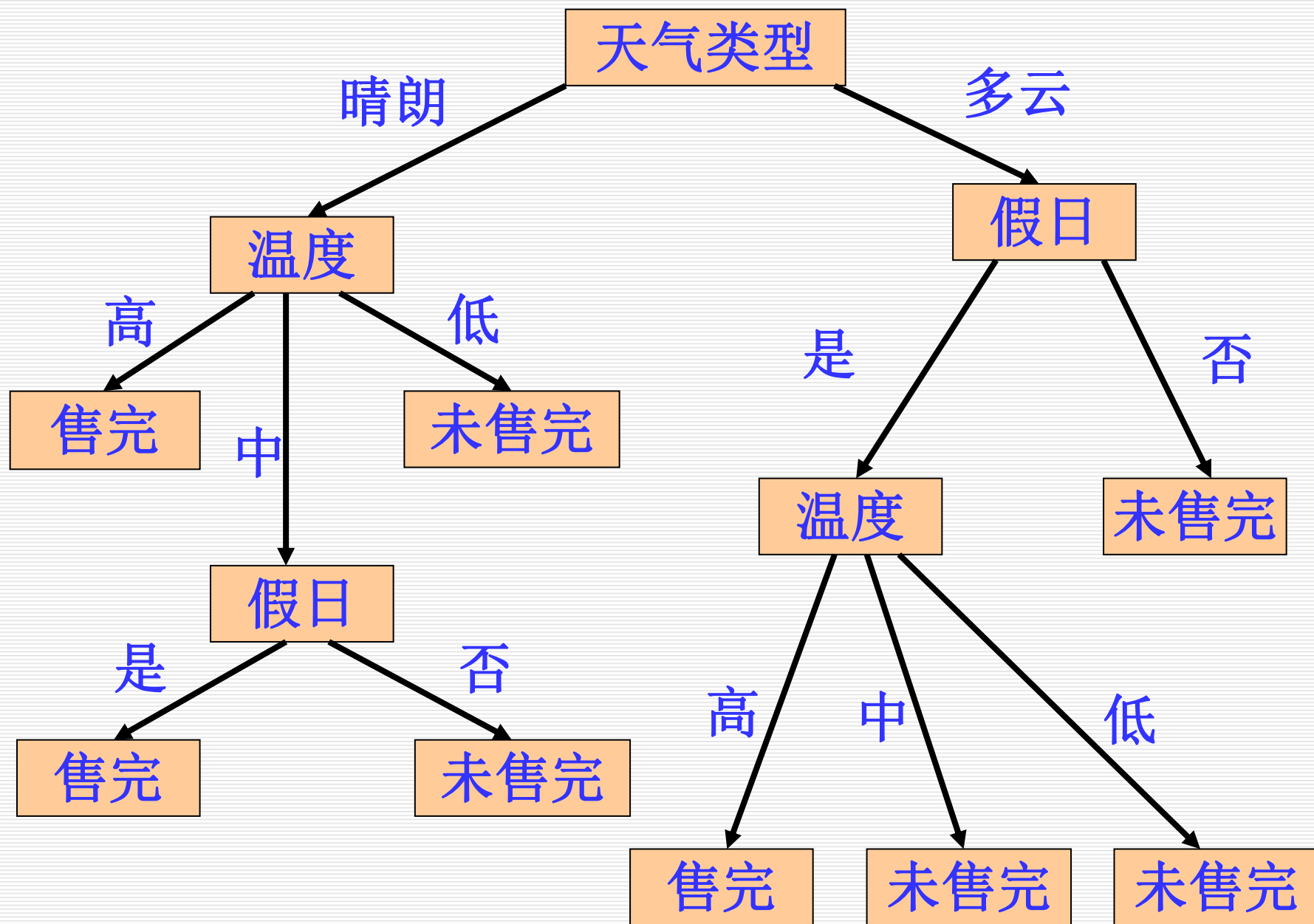
决策树

□ 决策树表示法

- 树的节点表示属性
- 树的分支表示属性值

□ 决策树通过把实例从根节点开始，最后排列到某个叶节点来进行分类，叶节点就是实例所属的类别

冰淇淋销售决策树



决策树

□ 决策树及其路径和分支

- 从根节点到叶节点的一条路径，对应一组属性值的合取式
- 树的各条分支，代表属性值一个析取式
- 决策树代表属性值合取后的析取式

3.1 销售决策树

□ 判断是否销售完成的决策树对应于以下表达式

$(\text{天气类型} = \text{晴朗} \wedge \text{温度} = \text{高}) \vee$

$(\text{天气类型} = \text{晴朗} \wedge \text{温度} = \text{中} \wedge \text{日期} = \text{假日}) \vee$

$(\text{天气类型} = \text{多云} \wedge \text{日期} = \text{假日} \wedge \text{温度} = \text{高})$

□ 决策树也可以表示成规则的形式

3.1 销售决策树

□ 销售完成的决策树对应于以下规则集

■ IF 天气类型=晴朗AND温度=高 THEN 售完

■ IF 天气类型=晴朗AND温度=中AND 日期=假日
THEN 售完

■ IF 天气类型=多云AND 日期= 假日 AND温度=高
THEN 售完

3.1 销售决策树

- ❑ 冰淇淋销售的决策树是通过观察和记录售完冰淇淋所对应的天气条件而建立的，具有代表性
- ❑ 可以根据决策树和未来天气情况预测冰淇淋的销售情况

决策树的应用

- 啤酒厂新建酒吧的位置评估问题
- 对酒吧的位置，需要调查很多属性，包括：
 - 落在城市还是小镇：Y/N
 - 坐落区域类型：L/M/S/N
 - 交通条件：G/P/A
 - 其它
- 该啤酒厂拥有一个数据库，其中包括现有酒吧及其相应属性值。

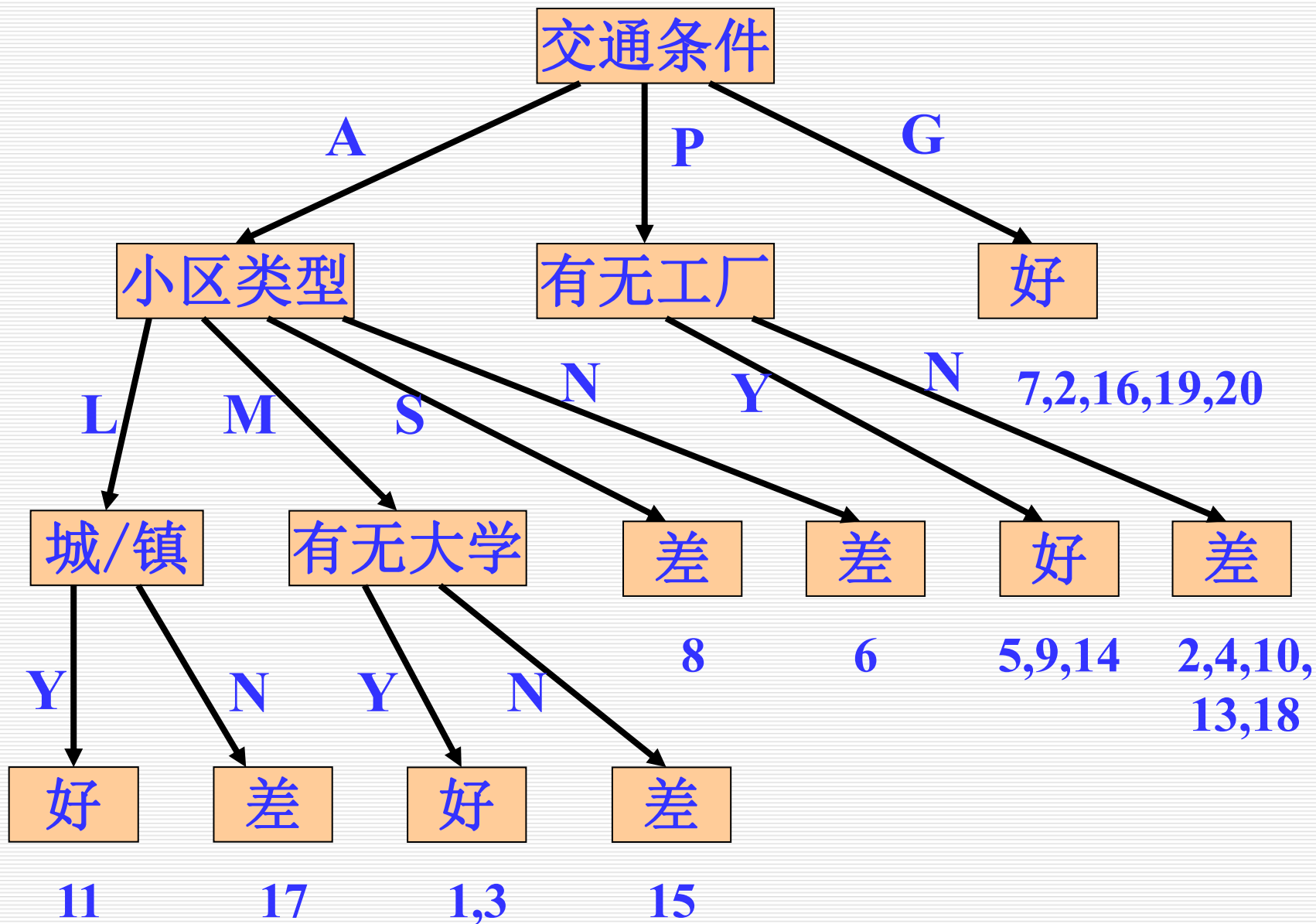
例子——酒吧位置

	城/镇	大学	小区	工厂	交通	数量	效益
1	Y	Y	M	N	A	L	G
2	N	N	S	N	P	L	P
3	Y	Y	M	N	A	M	G
4	Y	N	M	N	P	S	P
5	N	N	M	Y	P	M	G
6	N	Y	N	N	A	S	P
7	Y	N	N	N	G	S	G
8	Y	N	S	N	A	M	P
9	N	N	L	Y	P	L	G

例子——酒吧位置

- 说明：啤酒厂的酒吧有20个，表中只列出了9个酒吧的位置评估情况；全部的情形在后面的决策树标出
- 决策树给出了一些属性值规则的集合

酒吧位置 评估决策树



构造决策树的ID3算法

- 基本的ID3算法通过自顶向下构造决策树进行学习
 - 构造过程是从“哪一个属性将在树的根节点被测试”这个问题开始

构造决策树的ID3算法

□ ID3算法的思想

- 1. 根节点被赋予最佳的属性
- 2. 为根节点属性的取值进行分类
- 3. 将训练样例排列到适当的分支
- 4. 对子节点重复上述过程

构造决策树的ID3算法

□ 节点如何被赋予最佳的属性？

■ 熵

■ 信息增益

3.2 熵

(1) 系统混乱程度的度量---熵

假设一个系统 S 中含有两种信息，其中 p_1 表示第一种信息在 S 中的比例， p_2 表示第二种信息在 S 中的比例，那么系统 S 的熵为：

$$Entropy(S) = -p_1 \log_2 p_1 - p_2 \log_2 p_2$$

3.2 熵

例1. 如果S的所有成员属于同一类，那么：

$$Entropy(S) = -(1) \log_2 1 - 0 \log_2 0 = 0$$

例2. 如果S分别有一半成员属于一类，则：

$$Entropy(S) = -(0.5) \log_2 0.5 - 0.5 \log_2 0.5 = 1$$

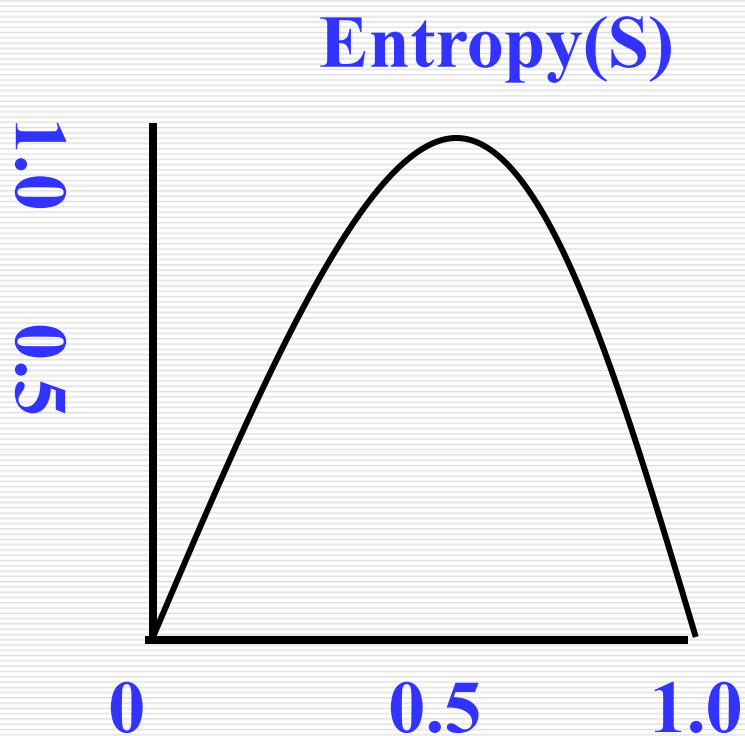
例3：假设S是一个有14个训练样例的集合，它包含9个正例和5个反例，并记为 $\{9+, 5-\}$ 来表示训练集。

则S的熵为：

$$\begin{aligned} Entropy(S) &= -(9/14) \log_2(9/14) - (5/14) \log_2(5/14) \\ &= 0.94 \end{aligned}$$

3.2 熵

□ 布尔分类时, 熵的值为 $[0,1]$, 熵曲线如下:



3.2 熵

(2)多种分类时，系统的熵

□ 在n种信息分类下，S的熵定义为：

$$Entropy(S) = \sum_{i=1}^n (-p_i \log_2 p_i)$$

其中， p_i 表示S中属于*i*类的比例；

$$Entropy(S) \in [0, \log_2 n]$$

3.2 熵

(3)根据属性，系统分类后的熵

设系统 S 的熵为 $Entropy(S)$, 且 S 具有属性 A , 可以用 $Values(A)$ 表示属性 A 所有可能值的集合, S_v 是 S 中属性 A 的值为 v 的集合, 则 S 分类后的期望熵为:

$$E_A(S) = \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

3.2 熵

例4: 假设 S 是一个有14个训练样例的集合, 它包含9个正例和5个反例, 则 $Entropy(S) = 0.94$; S 具有属性 $Wind$, 14个样例中具有 $Weak$ 值的有6个正例, 2个反例; 具有 $Strong$ 值的有3个正例, 3个反例; 则按照属性 $Wind$ 分类后, S 的熵为:

$$\begin{aligned} \sum_{v \in \{Strong, Weak\}} \frac{|S_v|}{S} Entropy(S_v) &= \\ (8/14) Entropy(S_{weak}) + (6/14) Entropy(S_{strong}) &= \\ = (8/14) [-(6/8) \log_2(6/8) - (2/8) \log_2(2/8)] + \\ (6/14) [-(3/6) \log_2(3/6) - (3/6) \log_2(3/6)] &= \\ = (8/14) 0.811 + (6/14) 1 = 0.892 \end{aligned}$$

3.2 熵

例5: 假设 S 是一个有14个训练样例的集合, 它包含9个正例和5个反例, 则 $Entropy(S) = 0.94$; S 具有属性 $Humidity$, 14个样例中具有 $High$ 值的有3个正例, 4个反例; 具有 $Normal$ 值的有6个正例, 1个反例。则按照属性 $Humidity$ 分类后, S 的熵为:

$$\begin{aligned} \sum_{v \in \{High, Normal\}} \frac{|S_v|}{S} Entropy(S_v) &= \\ (7/14) Entropy(S_{High}) + (7/14) Entropy(S_{Normal}) &= \\ = (7/14)[(-3/7) \log_2(3/7) - (4/7) \log_2(4/7)] + & \\ (7/14)[(-6/7) \log_2(6/7) - (1/7) \log_2(1/7)] & \\ = 0.789 \end{aligned}$$

3.3 最佳属性选择

□ 信息增益

—用于衡量某一个属性分类训练数据的能力

—一个属性**A**相对训练样例集合**S**的信息增益**Gain(S,A)**

$$Gain(S, A) \equiv Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

3.3 最佳属性选择

- **信息增益：**表示根据属性A分类而导致S的熵的减小值
- 在前面的例子中，根据风力和湿度分类的信息增益如下

$$Gain(S, Humidity) = 0.94 - 0.789 = 0.151$$

$$Gain(S, Wind) = 0.94 - 0.892 = 0.048$$

3.3 最佳属性选择

- 最佳属性就是使集合的信息增益最大的属性
 - 也就是使得集合分类后的熵值降到最低的那个属性

练习1

□ 假设训练样例集合S，共有6个样例，其中3个正例和3个反例。有两个属性A和B，计算：

—S相对于属性A和B的信息增益

(附： $\log_2 3 = 1.585$, $\log_2 5 = 2.322$)

练习1

实例	分类	A	B
1	正例	T	T
2	正例	T	T
3	反例	T	F
4	正例	F	F
5	反例	F	T
6	反例	F	T

练习1——答案

$$E(S) = 1 \quad (\text{附: } \log_2 3 = 1.585, \log_2 5 = 2.322)$$

$$\begin{aligned} E_A(S) &= \sum_{v \in \{T, F\}} \frac{|S_v|}{S} \text{Entropy}(S_v) \\ &= (3/6) \text{Entropy}(S_T) + (3/6) \text{Entropy}(S_F) \\ &= (3/6)[(-2/3) \log_2(2/3) - (1/3) \log_2(1/3)] + \\ &\quad (3/6)[(-1/3) \log_2(1/3) - (2/3) \log_2(2/3)] \\ &= (-2/3) \log_2(2/3) - (1/3) \log_2(1/3) \\ &= (-2/3)[1 - \log_2(3)] + (1/3) \log_2(3) \\ &= 0.918 \end{aligned}$$

练习1——答案

$$E_B(S) = \sum_{v \in \{T, F\}} \frac{|S_v|}{S} \text{Entropy}(S_v)$$

$$= (4 / 6) \text{Entropy}(S_T) + (2 / 6) \text{Entropy}(S_F)$$

$$= (4 / 6) \times 1 + (2 / 6) \times 1$$

$$= 1$$

$$G(S, A) = 1 - 0.918 = 0.082$$

$$G(S, B) = 1 - 1 = 0$$

练习2

□ 画出对以下样本进行分类的决策树

实例	属性X	属性Y	类别
1	T	F	正例
2	F	F	反例
3	T	F	正例
4	F	F	反例
5	T	F	正例
6	F	T	正例

练习2——答案

$$E(S) = -(4/6)\log_2(4/6) - (2/6)\log_2(2/6)$$

$$= (-2/3)[1 - \log_2(3)] + (1/3)\log_2(3)$$

$$= 0.918$$

(附: $\log_2 3 = 1.585$, $\log_2 5 = 2.322$)

$$E_X(S) = \sum_{v \in \{T, F\}} \frac{|S_v|}{S} \text{Entropy}(S_v)$$

$$= (3/6)\text{Entropy}(S_T) + (3/6)\text{Entropy}(S_F)$$

$$= (3/6) \times 0 + (3/6)[(-1/3)\log_2(1/3) - (2/3)\log_2(2/3)]$$

$$= (1/2) \times 0.918 = 0.459$$

$$E_Y(S) = \sum_{v \in \{T, F\}} \frac{|S_v|}{S} \text{Entropy}(S_v)$$

$$= (1/6)\text{Entropy}(S_T) + (5/6)\text{Entropy}(S_F)$$

$$= (1/6) \times 0 + (5/6)[(-3/5)\log_2(3/5) - (2/5)\log_2(2/5)]$$

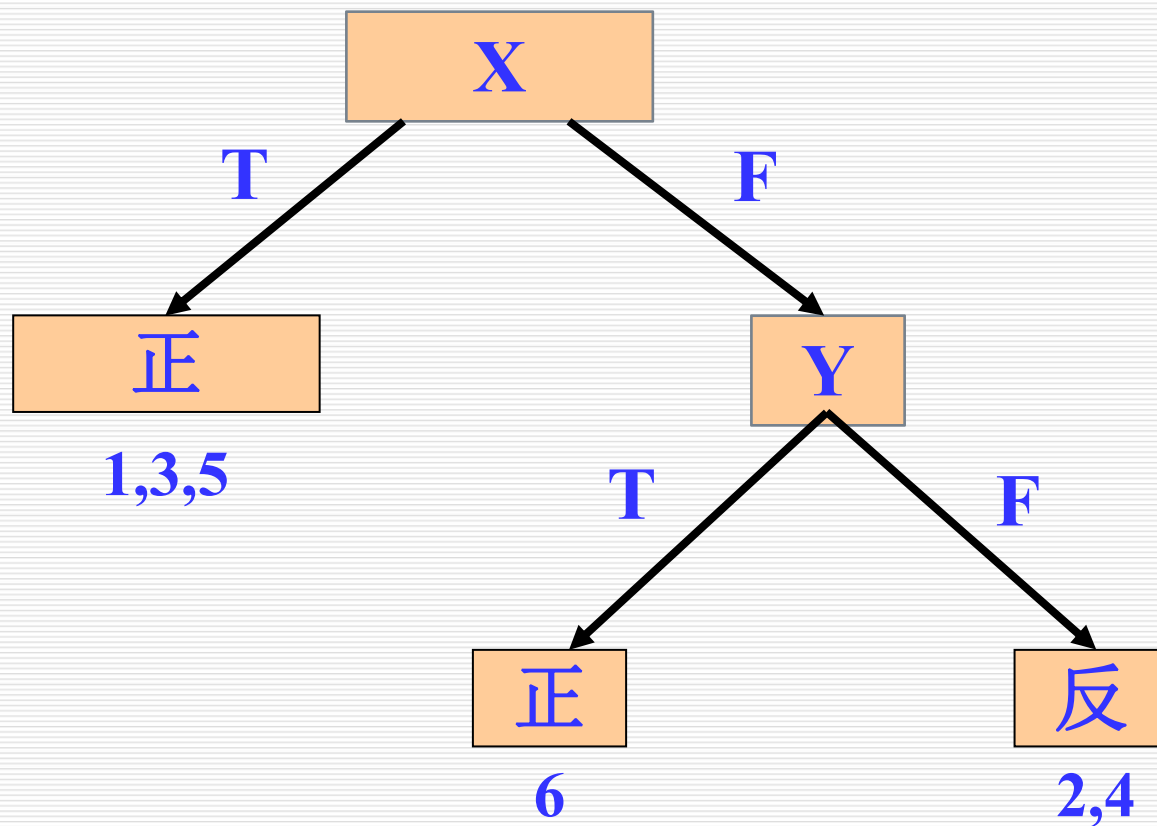
$$= (1/6)[(-3)(\log_2 3 - \log_2 5) - (2)(1 - \log_2 5)]$$

$$= 0.809$$

练习2——答案

$$G(S, X) = 0.918 - 0.459 = 0.459$$

$$G(S, Y) = 0.918 - 0.809 = 0.109$$



决策树学习的过度拟合问题

□ 手写体汉字识别

- 训练数据使用张三的笔迹

- 测试采用李四的笔迹

□ 自动驾驶

- 训练数据使用高速公路的图像

- 测试采用山路

决策树学习的过度拟合问题

过度拟合:

给定一个假设 $h \in$ 假设集合 H , 如果存在其它假设 $h' \in H$, 使得在训练样例上 h 的错误率比 h' 小, 但在整个实例分布上 h' 的错误率比 h 小, 那么就说假设 h 过度拟合训练集合。

决策树学习的过度拟合问题

□ 过度拟合发生

- 数据中有噪声
- 训练样例的数量太少
- 训练样例的种类不全

□ 避免过度拟合的策略：规则后修剪法

- 允许决策树过度拟合数据，然后对这个树进行后修剪

规则后修剪法

- 1. 从训练样例集合推导出**决策树**，使树生长直到尽可能好地拟合训练数据，允许过度拟合发生；
- 2. 将决策树转化为**等价的规则集合**，方法是对于从根节点到叶节点的每一条路径，创建一条规则；

规则后修剪法

- 3. 通过删除任何能导致估计精度提高的前件，来修剪每一条规则；
- 4. 按照修剪过的规则的估计精度，对它们进行排序，并按照该顺序，应用这些规则分类后来的实例

4. 概念学习

- 按照Mitchell的观点，学习的过程就是在假设空间上的搜索。搜索到一个特殊假设，使其和训练样例有最好的拟合度
- 概念学习就是对于给定某一类别的若干正例和反例，从中获得该类别的一般定义。

4. 概念学习

□ 概念学习

- 从训练样例空间中推断一个布尔函数，该函数对于每个训练样例，若为正例则输出1，若为反例则输出0

□ 概念学习的目标函数 $C(x)$

- 和机器学习模型中的 $C(x)$ 一致，即对于正例输出1，反例输出0

4. 概念学习

- 学习的任务是在整个实例集合 X 上，确定与目标概念 $C(x)$ 相同的假设 h ；但是已知的只是样例空间 D ，因此归纳学习算法最多只能保证输出的假设 h ，在 D 上与 $C(x)$ 一致，即 $h(x)=C(x)$,当 x 属于 D

4. 概念学习

□ 归纳学习的假设

- 对于未知实例($X-D$ 空间中的实例), 最好的假设就是与训练数据最佳拟合的假设

4.1 概念学习的搜索空间

- 每个概念可以看作一个对象或事件集合

- 可以是更大集合中选取的子类

- 如哺乳动物类的食肉动物

- 或是在这个较大集合中定义的布尔函数

- 如 $\text{bird}(x)$, 当 x =鸟类则 $\text{bird}(x)=\text{true}$

实例空间的大小

实例集合 X 的任意一个实例 x 可表示为一个 n 维向量 $\langle x_1, x_2, \dots, x_n \rangle$, 向量的每一个元素都是一个属性的取值。假设元素 x_i 的

取值有 $m_i (i = 1, 2, \dots, n)$ 个, 那么 X 的大小为: $\prod_{i=1}^n m_i$

假设空间H的大小

假设空间引入了两个特殊值 \perp 和 Φ ：

\perp 表示本属性可接受任意的属性值；

Φ 表示不接受任何属性值。

含有 Φ 的假设只能有一个， $\langle \Phi, \Phi, \dots, \Phi \rangle$ 。

含有实例空间 X 中的属性值的假设中不能含有 Φ ，但是 \perp 可以出现在任意假设的属性值中。因此，假设空间 H 的大小为：

$$1 + \prod_{j=1}^n (m_{ij} + 1)$$

假设空间H的大小

假设实例有4个属性 A, B, C, D , 属性值的个数分别为3,2,2,2, 那么实例空间 X 包括 $3 \times 2 \times 2 \times 2 = 24$ 种不同的实例, 假设空间 Y 有 $1 + 4 \times 3 \times 3 \times 3 = 109$ 种假设。随着属性个数和属性值的个数增大, 假设空间 H 迅速扩大, 因此学习过程需要有效地搜索无限大的假设空间, 以找到最佳拟合训练数据的假设。

练习

- 已知实例空间 X ，含有3个属性，属性值的个数分别为2, 4, 3。计算：
 - (1) 假设空间 H 的大小；
 - (2) 增加一个属性 A , 属性值的个数为3, 则实例空间和假设空间 H 的大小将增加多少？
 - (3) 推广到一般情形，新增属性 A , 属性值的个数为 k , 则实例空间和假设空间 H 的大小将增加多少？

答案

$$(1) |H| = 1 + 3 \times 5 \times 4 = 61$$

$$(2) |X| = 2 \times 4 \times 3 = 24$$

$$|X'| = 2 \times 4 \times 3 \times 3 = 72$$

$$|H'| = 1 + 3 \times 5 \times 4 \times 4 = 241$$

$$\Delta |X| = 48$$

$$\Delta |H| = 180$$

$$(3) \Delta |X| = 2 \times 4 \times 3 \times k - 2 \times 4 \times 3 = 24(k - 1)$$

$$\Delta |H| = [1 + 3 \times 5 \times 4 \times (k + 1)] - (1 + 3 \times 5 \times 4) = 60k$$

4.2 假设空间的偏序关系

- 利用假设的从一般到特殊偏序关系，可以在无限的假设空间中进行彻底的搜索，而不需要明确地列举所有的假设。
- 偏序关系由 \supseteq 和 Φ 构成

4.2 假设空间的偏序关系

例： $h_1 = \langle \text{Sunny}, ?, ?, \text{Strong}, ?, ? \rangle$

$h_2 = \langle \text{Sunny}, ?, ?, ?, ?, ? \rangle$, 看看哪些实例

可被 h_1 和 h_2 划分为正例？

显然， h_2 包含的实例约束较少，任何被 h_1 划分成正例的实例都会被 h_2 划分为正例；
即 h_2 比 h_1 更一般

4.2 假设空间的偏序关系

偏序关系的定义：

令 h_j 和 h_k 为在 X 上定义的布尔函数。称 h_j 比 h_k 更一般，并标记为 $h_j \geq gh_k$ ，当且仅当：

$$(\forall x \in X)[(h_k(x) = 1) \rightarrow (h_j(x) = 1)]$$

显然，假设空间 H 上最一般的假设是 $\langle ?, ?, \dots, ? \rangle$ ，最特殊的假设为 $\langle \Phi, \Phi, \dots, \Phi \rangle$

4.2 假设空间的偏序关系

推论：假设 h_j 比 h_k 更一般，即 $h_j \geq gh_k$ ，则有

- (1) 若 h_j 将一个实例判定为反例，则 h_k 一定将一个实例判定为反例；
- (2) 若 h_k 将一个实例判定为正例，则 h_j 一定将一个实例判定为正例；
- (3) 若 h_k 将一个实例判定为反例，则 h_j 不一定将一个实例判定为反例；
- (4) 若 h_j 将一个实例判定为正例，则 h_k 不一定将一个实例判定为正例；

4.2 假设空间的偏序关系

□ 上述四个推论非常重要，构成了设计搜索算法的基础

■ 可以利用这种偏序结构，有效搜索假设空间

4.4 FIND-S算法

- FIND-S算法利用偏序关系，搜索一个特殊的假设
 - 使得对于每个训练样例，当其为正例时，该假设将其判为正例；
 - 当其为反例时，该假设将其判为反例。

4.4 FIND-S算法

□ FIND-S算法思路

- 从H中最特殊的假设出发，然后在该假设将正例错判为反例时，根据推论3，将其极小一般化，从而使该假设将正例判为正例

4.4 FIND-S算法

FIND – *S*算法

1. 将 h 初始化为 H 中的最特殊假设；
2. 对每个正例 x , 对 h 的每个属性约束 a_i , 如果 x 满足 a_i , 那么不做任何处理； 否则将 h 中 a_i 替换为 x 满足的另一个更一般的约束；
3. 输出假设 h 。

例 题

- 张三进行水上运动的日子。共有**6**个属性，四个训练样例，其中三个正例和一个反例。

例子	天气	气温	湿度	风力	水温	预测	适宜
1	晴	暖	正常	大	暖	相同	是
2	晴	暖	高	大	暖	相同	是
3	雨	冷	高	大	暖	不同	否
4	晴	暖	高	大	凉	不同	是

过程

□ 求解假设集合H中的h，使得对于D中任意的d, 有 $h(d)=C(d)$.

□ 应用FIND-S算法：

1.将h初始化为H中最特殊的假设：

$$h \leftarrow \langle \Phi, \Phi, \Phi, \Phi, \Phi, \Phi \rangle$$

2.考虑第一个样例，这是一个正例。因为h的每一个属性值显然都不满足该样例的约束，所以都被替换掉：

$h \leftarrow \langle \text{晴}, \text{暖}, \text{正常}, \text{大}, \text{暖}, \text{相同} \rangle$

实际上，每个概念学习任务第一次泛化后均为第一个正例本身。

过程

3.考虑第二个样例，同样为正例。用 h 代入，由于湿度不满足约束，所以判为反例；此时需要将湿度属性值一般化，即：

$h \leftarrow \langle \text{晴}, \text{暖}, ?, \text{大}, \text{暖}, \text{相同} \rangle$

过程

4.由于第三个样例是个反例，不要考虑。

5.再考虑第四个样例，这是个正例。该算法进一步将 h 泛化，使之将该训练样例判为正例，

$$h \leftarrow \langle \text{晴}, \text{暖}, ?, \text{大}, ?, ? \rangle$$

经过训练样例集合 D 的训练，机器学习的结果是：
在天晴、暖和与风力大的日子，适合做水上运动。

FIND-S算法的特点

- 对于以属性约束的合取式描述的假设空间，算法保证输出的是 H 中与所有正例一致的最特殊的假设。
- FIND-S算法虽然只针对样例集合中的正例进行处理，对于反例未进行任何处理，但是它一定会将反例判为反例。

正确判断反例的证明

- 1. 算法输出的假设 h ，可以将训练样例集合中所有的正例判定为正例；
- 2. 对理想的目标函数 $C(x)$ ， $C(x)$ 在整个实例空间内都能正确判断，对于更小的训练空间 D 内，它当然可以将所有的正例都判定为正例。即 h 判为正例， $C(x)$ 肯定判为正例，则 $C(x)$ 比 h 更一般；
- 3. 根据推论1，若 $C(x)$ 将训练样例判为反例，则 h 也一定将该样例判为反例

FIND-S算法的问题

- 学习过程是否收敛到了正确的目标概念？
 - FIND-S算法找到了与训练数据一致的假设，但是不能确定是否找到了唯一合适的假设，即目标概念；
- 为何要用最特殊的假设？
 - FIND-S算法找到的是与训练数据一致的假设中最特殊的一个，而从这个假设开始到最一般的假设，都可以作为学习结果

FIND-S算法的问题

□ 训练样例是否一致？

- 在多数实际的学习问题中，训练数据常带有噪声和错误，这将严重破坏FIND-S算法。
- 因为算法忽略反例情形，如本来是正例因为错误变成反例，则导致算法结果出错。

练习1

□ 试利用**FIND-S**算法，获取对应下列训练集合的假设，并详细列出计算**h**的每个步骤。

例子	房子	宝马	公司	工厂	年龄	分类
1	别墅	否	否	大	大	正例
2	公寓	否	是	无	中	反例
3	别墅	是	否	大	中	正例
4	别墅	是	是	无	中	正例

练习1——答案

$$1) \ h_1 = \langle \Phi_1, \Phi_2, \Phi_3, \Phi_4, \Phi_5 \rangle$$

$$2) \ h_2 = \langle \text{別墅}, \text{否}, \text{否}, \text{大}, \text{大} \rangle$$

$$3) \ h_3 = \langle \text{別墅}, ?, \text{否}, \text{大}, ? \rangle$$

$$4) \ h_3 = \langle \text{別墅}, ?, ?, ?, ? \rangle$$

练习2

□ 利用**FIND-S**算法，计算对应下列训练集合的假设，并详细列出计算**h**的每个步骤。

例子	学历	身高	性格	家境	收入	分类
1	高	矮	好	穷	高	反例
2	中	高	差	富	高	正例
3	高	高	好	穷	高	正例
4	中	矮	好	富	中	反例
5	中	高	差	富	高	正例

练习2

□ 如果训练样例5修改成下表，利用**FIND-S**算法重新计算h

例子	学历	身高	性格	家境	收入	分类
1	高	矮	好	穷	高	反例
2	中	高	差	富	高	正例
3	高	高	好	穷	高	正例
4	中	矮	好	富	中	反例
5	中	中	差	富	高	正例

小 结

- 机器学习的定义
- 机器学习的模型
- 决策树学习
- 概念学习

附参考教材内容7.5 机器学习

7.5.1 机器学习的基本概念

- **机器学习**：机器学习（**Machine learning**）使计算机能模拟人的学习行为，自动地通过学习来获取知识和技能，不断改善性能，实现自我完善。

机器学习主要研究以下三方面问题：

- （1）**学习机理**：人类获取知识、技能和抽象概念的天赋能力。
- （2）**学习方法**：机器学习方法的构造是在对生物学习机理进行简化的基础上，用计算的方法进行再现。
- （3）**学习系统**：能够在一定程度上实现机器学习的系统。

7.5.1 机器学习的基本概念

- 一个学习系统一般应该有环境、学习、知识库、执行与评价等四个基本部分组成。

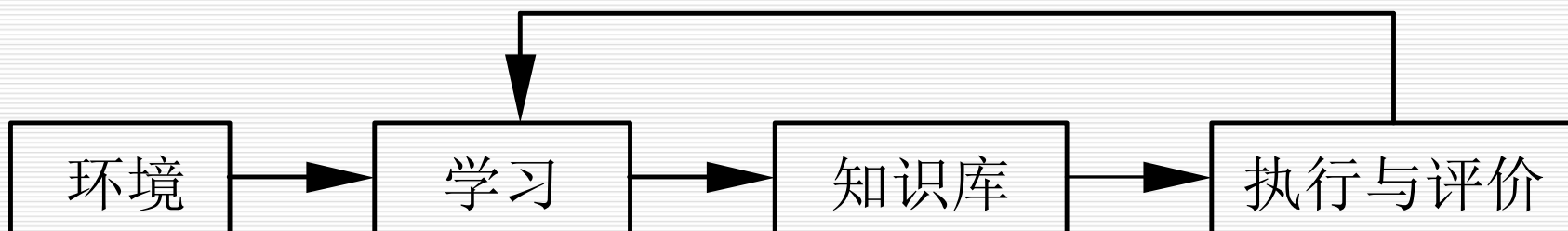


图7-1 学习系统的基本结构

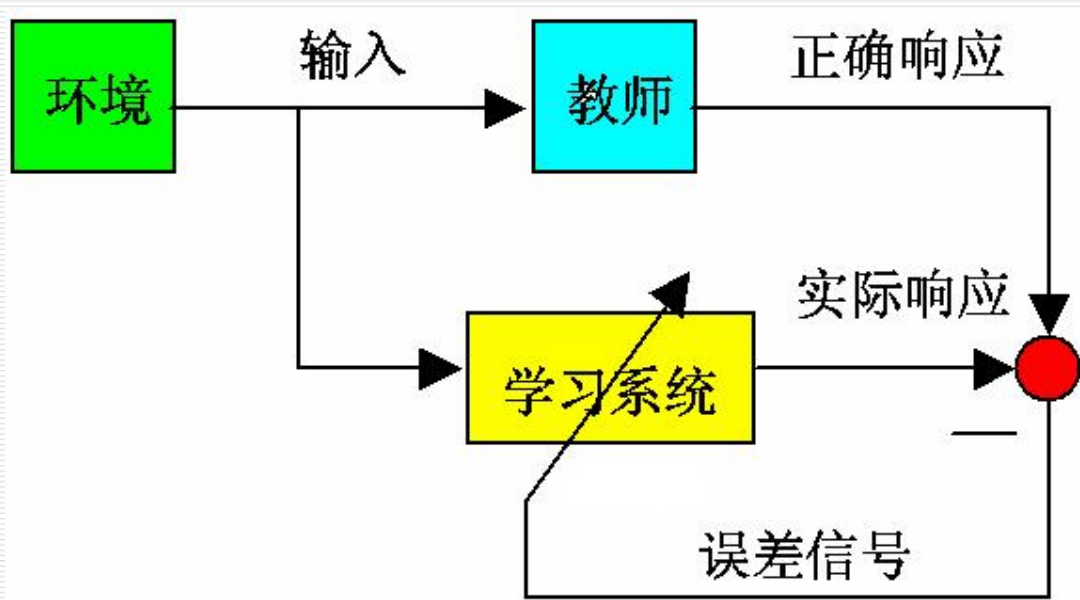
7.5.2 机器学习的分类

1. 按学习方法分类（温斯顿，1977）

- 机械式学习、指导式学习、示例学习、类比学习、解释学习等。

2. 按学习能力分类：

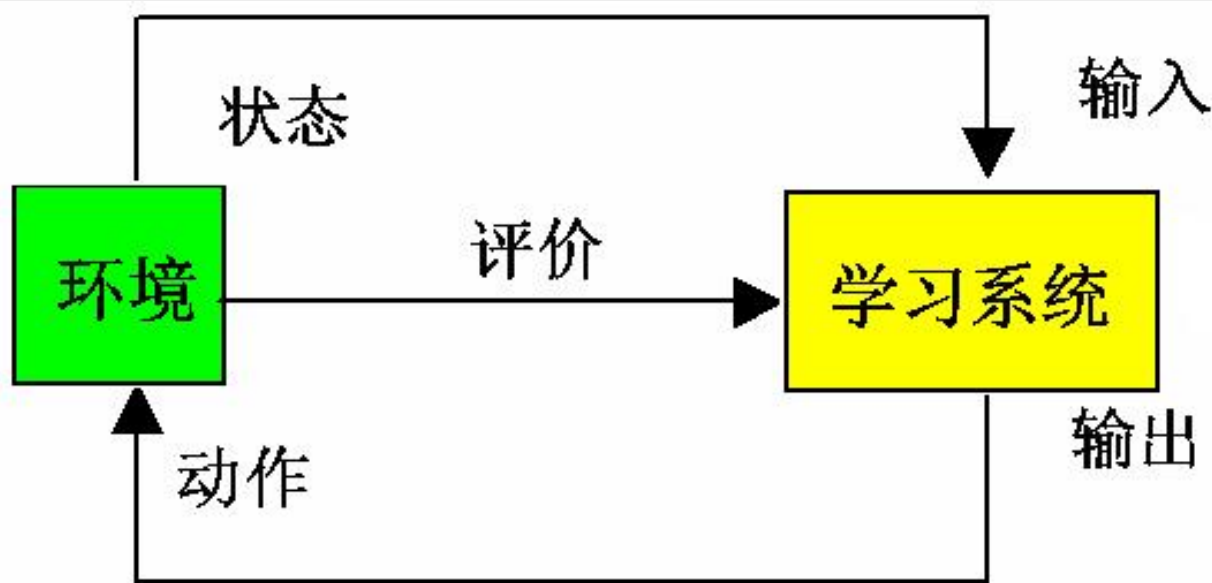
- 监督学习（有教师学习）



7.5.2 机器学习的分类

2. 按学习能力分类：

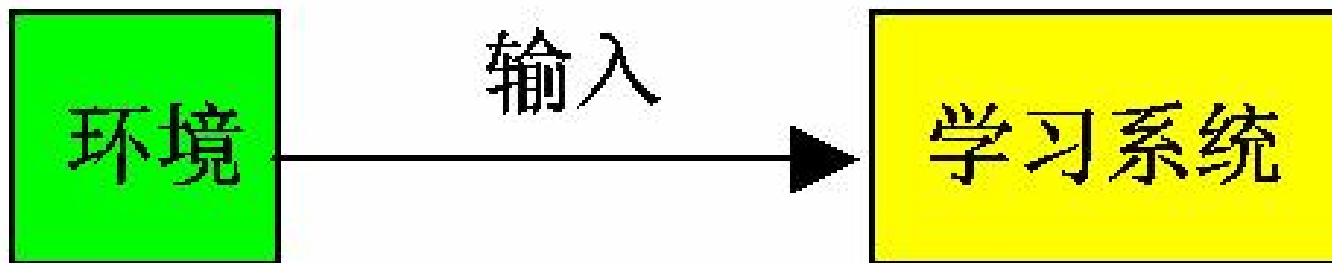
- 强化学习（再励学习或增强学习）



7.5.2 机器学习的分类

2. 按学习能力分类:

- 非监督学习（无教师学习）



7.5.2 机器学习的分类

3. 按推理方式分类:

- 基于演绎的学习（解释学习）。
- 基于归纳的学习（示例学习、发现学习等）。

4. 按综合属性分类:

- 归纳学习、分析学习、连接学习、 遗传式学习等。

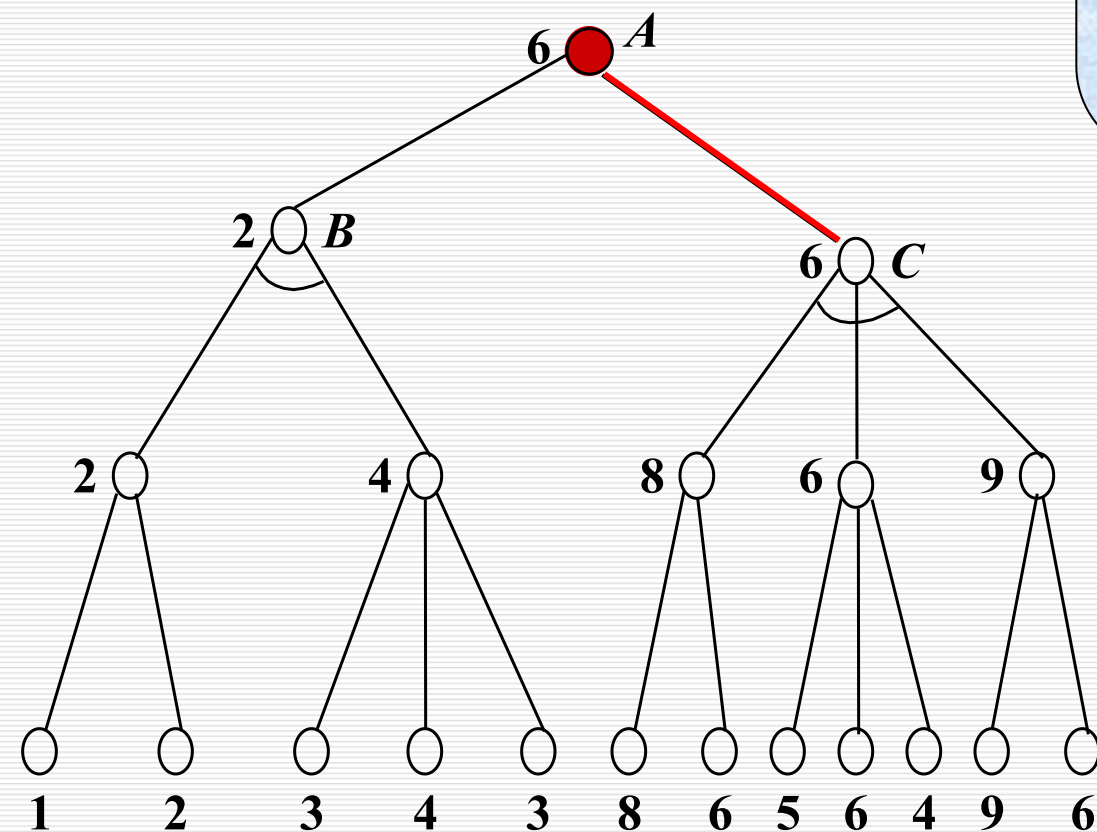
7.5.3 机械式学习

- 机械式学习（rote learning）又称记忆学习，或死记式学习：通过直接记忆或者存储外部环境所提供的信息达到学习的目的，并在以后通过对知识库的检索得到相应的知识直接用来求解问题。
- 机械式学习实质是用存储空间来换取处理时间。
- 典型例子：1959年，塞缪尔(A. L. Samuel)的跳棋程序CHECKERS。

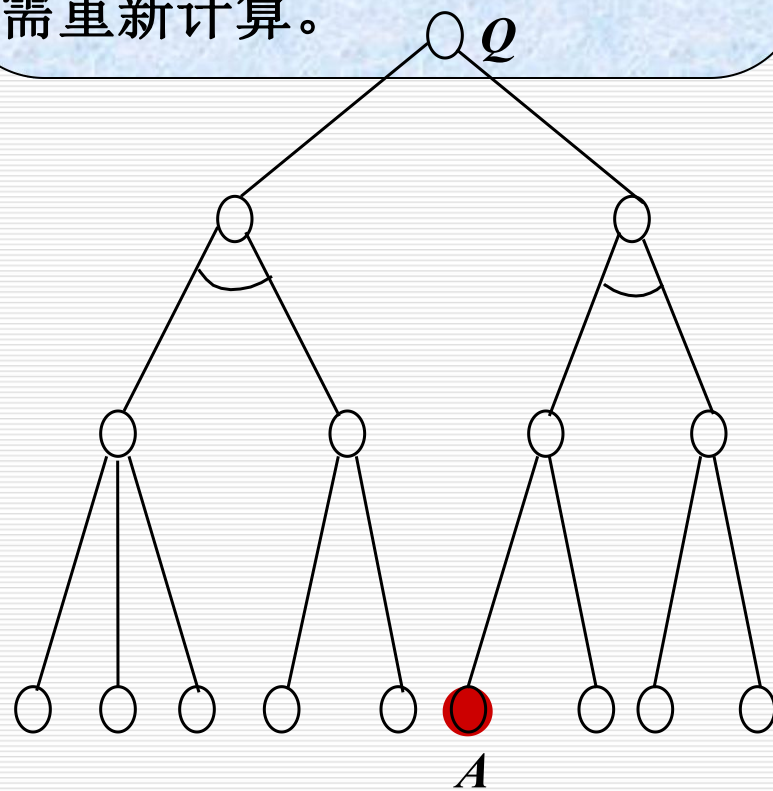
7.5.3 机械式学习

□ 塞缪尔的跳棋程序 CHECKERS

- 在给定搜索深度下用估价函数对格局进行评分，通过倒推计算求出上层节点的倒推值，决定当前的最佳走步。
- 下次遇到相同情况，直接利用倒推值决定最佳走步，不需重新计算。



博弈搜索树



以 A 为结点的博弈树

7.5.3 机械式学习

□ **机械式学习（rote learning）** 又称记忆学习，或死记式学习：通过直接记忆或者**存储**外部环境所提供的信息达到学习的目的，并在以后通过对知识库的**检索**得到相应的知识直接用来求解问题。

□ 机械学习的主要问题：

- **存储组织信息**：要采用适当的存储方式，使检索速度尽可能地快。
- **环境的稳定性与存储信息的适用性问题**：机械学习系统必须保证所保存的信息适应于外界环境变化的需要。
- **存储与计算之间的权衡**：对于机械学习来说很重要的一点是它不能降低系统的效率。

7.5 机器学习

- 7.5.1 机器学习的基本概念
- 7.5.2 机器学习的分类
- 7.5.3 机械式学习
- 7.5.4 指导式学习
- 7.5.5 示例学习

7.5.4 指导式学习

□ **指导式学习**（**learning by being told**）又称**嘱咐式学习**或**教授式学习**：由外部环境向系统提供一般性的指示或建议，系统把它们具体地转化为细节知识并送入知识库中。在学习过程中要反复对形成的知识进行评价，使其不断完善。

□ **指导式学习的学习过程**：征询指导者的指示或建议、把征询意见转换为可执行的内部形式、加入知识库、评价。

7.5.4 指导式学习

1. 征询指导者的指示或建议

- **简单征询：**指导者给出一般性的意见，系统将其具体化。
- **复杂征询：**系统不仅要求指导者给出一般性的建议，而且还要具体地鉴别知识库中可能存在的问题，并给出修改意见。
- **被动征询：**系统只是被动地等待指导者提供意见。
- **主动征询：**系统不只是被动地接受指示，而且还能主动地提出询问，把指导者的注意力集中在特定的问题上。

7.5.4 指导式学习

2. 把征询意见转换为可执行的内部形式

□ 学习系统应具有把用约定形式表示的征询意见转化为计算机内部可执行形式的能力，并且能在转化过程中进行语法检查及适当的语义分析。

3. 加入知识库

■ 在加入过程中要对知识进行一致性检查，以防止出现矛盾、冗余、环路等问题。

4. 评价

■ 评价方法：对新知识进行经验测试，即执行一些标准例子，然后检查执行情况是否与已知情况一致。

7.5 机器学习

- 7.5.1 机器学习的基本概念
- 7.5.2 机器学习的分类
- 7.5.3 机械式学习
- 7.5.4 指导式学习
- 7.5.5 示例学习

7.5.5 示例学习

□ **示例学习**（**learning from examples**，**实例学习**或**从例子中学习**）：通过从环境中取得若干与某概念有关的例子，经归纳得出一般性概念的一种学习方法。

□ 示例学习中，外部环境（教师）提供一组例子（**正例和反例**），然后从这些特殊知识中归纳出**适用于更大范围的一般性知识**，它将**覆盖所有的正例并排除所有反例**。



7.5.5 示例学习

示例学习的学习模型

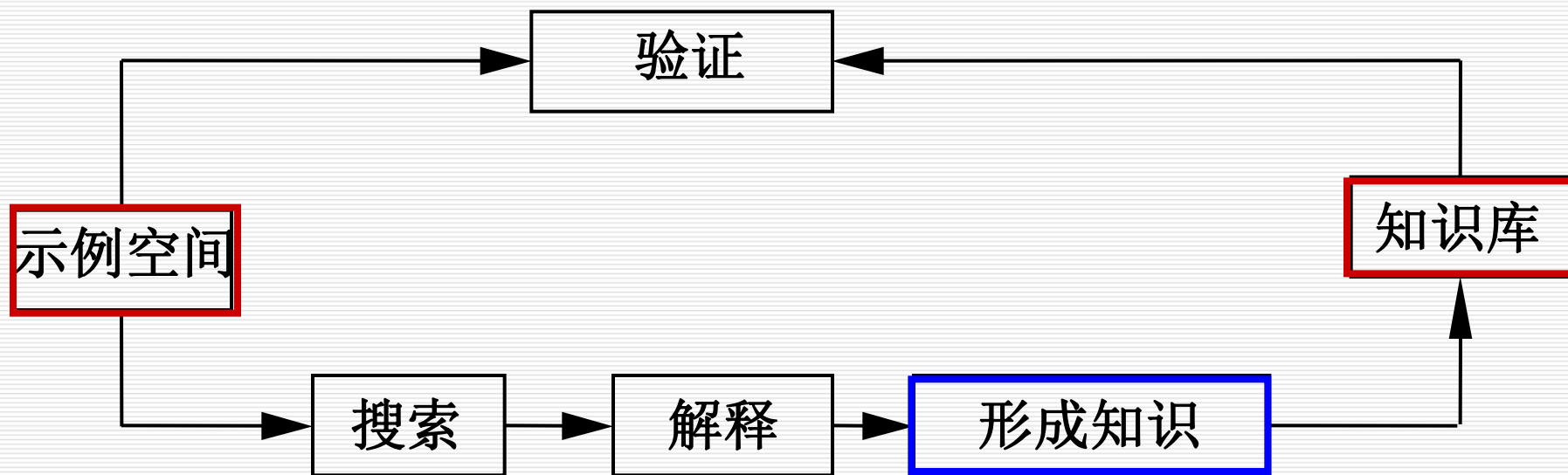


图7.7 示例学习的学习模型