

# 第9讲 HW

万晨阳 3210105327

## 2.1

1. 假设空间的大小是指假设空间中每个属性的可能值的不同组合数。假设空间中每个属性的值除了"?" (可接受任何值) 和 " $\emptyset$ " (不可接受任何值) 外, 还包括可分配的具体值。由于任何属性的 " $\emptyset$ " 符号都会将实例归类为负值, 因此我们只计算一次带有该值的所有假设。将其视为 "从不" 假设。各个属性的value数如下:

attribute	value
sky	3 + 1
air temp	2 + 1
humidity	2 + 1
wind	2 + 1
water	2 + 1
forecast	2 + 1

实例数为  $3 \times 2 \times 2 \times 2 \times 2 \times 2 = 96$ 。

而对于假设数, 后面的+1是因为考虑了每个属性都可以取"?" (可接受任意值)。这样总的假设空间大小为:

$4 \times 3 \times 3 \times 3 \times 3 \times 3 = 972$ , 再加上最特殊的  $\langle \emptyset, \emptyset, \emptyset, \emptyset, \emptyset, \emptyset \rangle$  (每个都是反例) 的情况, 总共为  $972 + 1 = 973$ 。也就是假设空间的大小是973。

2. 加上这一个新的属性之后, 表中多了一项。所以可能的实例数

attribute	value
water current	3 + 1

实例数:  $3 \times 3 \times 2 \times 2 \times 2 \times 2 \times 2 = 288$

增加的实例数:  $288 - 96 = 192$

假设数:  $4 \times 4 \times 3 \times 3 \times 3 \times 3 \times 3 + 1 = 3889$

增加的假设数:  $3889 - 973 = 2916$

即实例数增加了192, 假设数增加了2916。

3. 推广到更加一般的情况, 新增加的属性A有k种取值:

attribute	value
A	k + 1

则实例数变为 当前实例数  $\times k$ , 增加的实例数为 当前实例数  $\times (k - 1)$

假设数变为 (当前假设数  $\times (k + 1) - k$ , 增加的假设数为 当前假设数  $\times k - k$

对于该问题, 新增加的属性A有k种取值时, 实例数变为  $96 \times k$ , 增加的实例数为  $96 \times (k - 1)$

假设数变为  $(973 \times (k + 1)) - k$ , 增加的假设数为  $972 \times k$

## 2.2

Example	Sky	AirTemp	Humidity	Wind	Water	Forecast	EnjoySport
1	Sunny	Warm	Normal	Strong	Warm	Same	Yes
2	Sunny	Warm	High	Strong	Warm	Same	Yes
3	Rainy	Cold	High	Strong	Warm	Change	No
4	Sunny	Warm	High	Strong	Cool	Change	Yes

1. 分布给出S和G边界集合：

- 初始化

$$S_0 = \{ \langle \emptyset, \emptyset, \emptyset, \emptyset, \emptyset, \emptyset \rangle \}$$

$$G_0 = \{ \langle ?, ?, ?, ?, ?, ? \rangle \}$$

- 训练例4（正例）

$$S_1 = \{ \langle \text{Sunny}, \text{Warm}, \text{High}, \text{Strong}, \text{Cool}, \text{Change} \rangle \}$$

$$G_1 = \{ \langle ?, ?, ?, ?, ?, ? \rangle \}$$

- 训练例3（反例）

$$S_2 = \{ \langle \text{Sunny}, \text{Warm}, \text{High}, \text{Strong}, \text{Cool}, \text{Change} \rangle \}$$

$$G_2 = \{ \langle \text{Sunny}, ?, ?, ?, ?, ? \rangle, \langle ?, \text{Warm}, ?, ?, ?, ? \rangle, \langle ?, ?, ?, ?, \text{Cool}, ? \rangle \}$$

- 训练例2（正例）

$$S_3 = \{ \langle \text{Sunny}, \text{Warm}, \text{High}, \text{Strong}, ?, ? \rangle \}$$

$$G_3 = \{ \langle \text{Sunny}, ?, ?, ?, ?, ? \rangle, \langle ?, \text{Warm}, ?, ?, ?, ? \rangle \}$$

- 训练例1（正例）

$$S_4 = \{ \langle \text{Sunny}, \text{Warm}, ?, \text{Strong}, ?, ? \rangle \}$$

$$G_4 = \{ \langle \text{Sunny}, ?, ?, ?, ?, ? \rangle, \langle ?, \text{Warm}, ?, ?, ?, ? \rangle \}$$

2. 对于我们的示例集合大小，最小值为  $2 \times (4 + 1) = 10$ 。由于使用的假设表示法只能用"?"来概括一个属性，因此 S 中在任何时候都只能有一个假设。由于特定的假设空间依赖于连接表达式，因此假设永远不会被添加到 S 中。换句话说，不能使用"或"条件来表达S。所以为了达到最小化的目标，我们应该关注 G 中的假设数。我们注意到，如前面训练过程所述，只有在出现反例时，才会将假设添加到 G 中。**因此，一种一般的策略是在整个训练过程中先呈现所有正例，然后再呈现反例。**通过这一策略，我们最终得到的集合大小为11，是这组训练示例的最小值。（因为最终状态中 G 包含 2 个假设，比上面计算的绝对最小值10多一个）

遵循以上原则（先呈现所有正例，然后再呈现反例），一种能达到题目要求的可能样例排序为：样例1、样例2、样例4、样例3。我们最终得到所有S和G的集合中间结果的大小之和为11。