6 ccs3ain,教程 03 (版本 1.0)

这些问题中有几个利用了图 1 中的贝叶斯网络。

1.写下一个贝叶斯网络来捕捉以下知识:(a)吸烟会致癌;(b)吸烟会导致口臭。鉴于以下信息:

P(癌症|吸烟)= 0.6 P(口臭|吸烟)= 0.95

P(吸烟)= 0.2，使用朴素贝叶斯模型确定 P(吸烟、癌症、口臭)。

假设这三个变量都是二元的:吸烟有“吸烟”和“吸烟”的值，癌症有“癌症”和“癌症”的值，口臭有“口臭”和“口臭”的值。

2.写一个贝叶斯网络来捕捉以下知识:(a)晚开始会导致学生的项目失败，(b)忽视建议会导致学生的项目失败。使用二进制变量 LateStart，值为 l 和 l, IgnoreAdvice，值为 i 和 i, FailProject，值为 f 和 f。

你知道 LateStart 和 IgnoreAdvice 是失败项目的非交互原因，所以使用 noise - or 模型来构建与三个变量相关的条件概率表:

$$P(f|l) = 0.7$$

$$P(f|i) = 0.8$$

3.在图 1 中，我们需要多少个数字来指定所有必要的网络概率值?如果没有条件独立性(如果我们没有网络)，我们需要多少个?

4.计算联合概率 P(m, t, h, s, c)

5.使用枚举算法计算 P(m|h, s)的概率。

6.使用先验抽样来计算 m, t, h, s 和 c 的联合概率。

使用表 1 中的随机数序列。只给出前 5 个样本的结果。

如果您需要比表 1 中存在的随机数更多的随机数，那么重新开始该序列。

7.采用拒绝采样计算 P(m|h, s)

使用表 1 中的随机数序列，从序列的开头开始。这次你应该展示前 5 个没有被拒绝的样品的结果。

如果您需要比表 1 中存在的随机数更多的随机数，那么重新开始该序列。

8.用重要性抽样法计算 P(m|h, s)。

使用表 1 中的随机数序列，从序列的开头开始，然后再次报告前五个样本的结果。

如果您需要比表 1 中存在的随机数更多的随机数，那么重新开始该序列。

9.使用吉布斯抽样计算 P(m|h, s)。

使用表 1 中的随机数序列，只给出前 5 个样本的结果。

如果您需要比表 1 中存在的随机数更多的随机数，那么重新开始该序列。10.对于本教程的可选计算部分，请从 KEATS 下载文件 wetGrass.py。这提供了使用石榴的代码[1]，它实现了 Lecture/Week 3 中的“Wet Grass”示例。你可以使用以下命令运行这个示例:

1 关于石榴的详细信息请参见教程 2。

python wetGrass.py

这应该会给你这个输出:

证据是:草是湿的。

这是预测:
多云， 洒水， 下雨， 湿草。
通过数字:
阴天:(:(c, 0.6010250059300832)， (l, 0.3989749940699168)洒水:((f, 0.5976922222293125)， (n, 0.40230777777068755)雨水:(((r, 0.7772459441092383)， (n, 0.2227540558907617))

py 的第一部分设置贝叶斯网络。这是内衣。py 的简单扩展。

然后，第二部分在模型上运行查询。变量的场景:

scenario =[[无，无，无，' w ']]

指定模型中观察到的变量值(证据)——变量的顺序是 model.add_states()命令中指定的顺序。在这里，我们将变量 WetGrass 设置为 w 值，表示草是湿的，而没有说任何关于多云、洒水器和雨的值。这样的:

证据是:草是湿的。

反映输入的反码做这在某种程度上是通用的方式 [2].

该代码使用两种方法根据证据进行推理。第一个 model.predict()确定模型中每个变量最有可能(最高概率)的值:

这是预测:
多云， 洒水， 下雨， 湿草。

然后使用 model.predict_proba()来生成所有非证据变量值的概率:

通过数字:
阴天:(:(c, 0.6010250059300832)， (l, 0.3989749940699168)洒水:((f, 0.5976922222293125)， (n, 0.40230777777068755)雨水:(((r, 0.7772459441092383)， (n, 0.2227540558907617))

结果表明 model.predict_proba()处理证据变量的方式与处理非证据变量的方式不同，因此代码不会输出任何关于证据变量的信息(毕竟我们已经知道证据变量所取的值)。

代码的最后一部分对 model.predict()和 model.predict_proba()生成的内容进行了少量的格式化，以生成上面的输出。

(a)使用 wetGrass.py 计算模型中变量的概率，有:

草地湿了，洒水器开着。

2 洒水喷头开着，天空多云。

3 下雨了。

(b)用石榴构建图 1 中的模型。用该模型计算 P(m|h, s)。

2 .石榴的文档不是很好，但是我可以告诉你，一旦模型建立，没有函数可以访问模型中的状态，所以我们不能以一般的方式检索变量，以及它们的名称和值。(当然，我们可以建立自己的结构来记录这一点)。因此它们是硬编码的。

0. 0. 0. 0.43 0.96 0. 0.89 0. 0.34

表 1:一些 0 和 1 之间的随机数。

0.1 P
(M)

米	P (S | M)
T	0
F	0

米	P (T | M)
T	0
F	0

年	T	P(c | s, t)
T	T	0
T	F	0.85
F	T	0.85
F	F	0.

T	P (H | T)
T	0
F	0

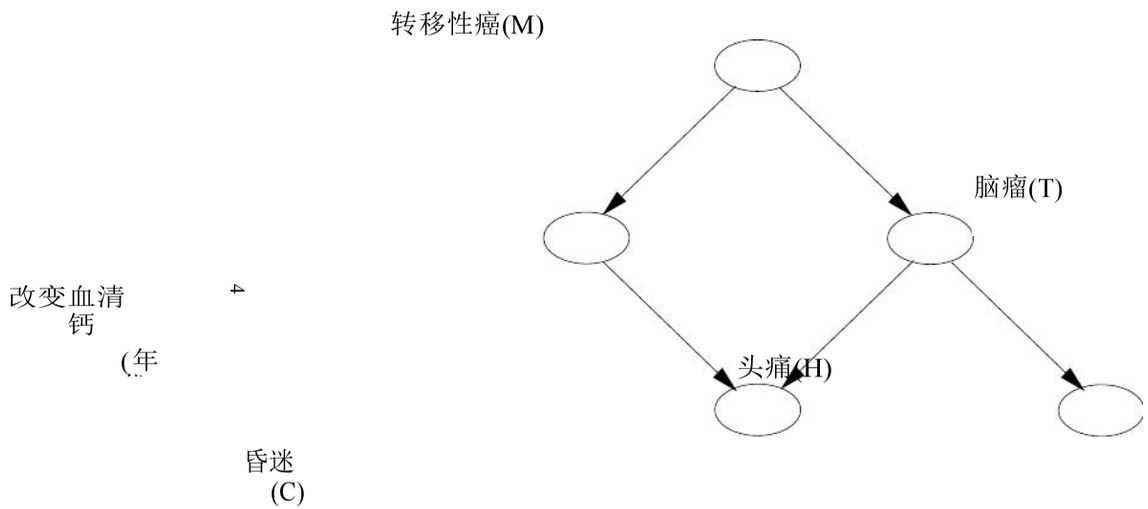
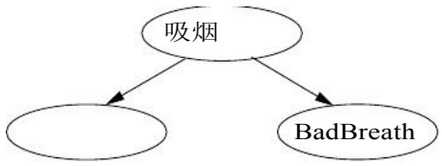


图 1:贝叶斯网络示例。

第 03 课回答

(版本 1.0)

1.贝叶斯网络为:



告知:

Psmoking) = 0.2

P(癌症|吸烟)= 0.6

P (badBreath |吸烟)= 0.95

现在，朴素贝叶斯模型告诉我们:

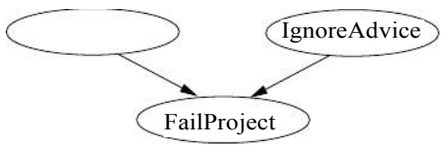
P(原因，结果 1，…， P(Cause)Y P(Effect i|Cause) i

所以:

P(吸烟、癌症、口臭)= P(吸烟) • P(癌症|吸烟) • P(口臭|吸烟)= 0.2 • 0.6 • 0.95

= 0.114

2.贝叶斯网络是:LateStart



3.在没有网络的情况下，需要 5 个变量:

$$2^5 - 1 = 31$$

数字。有了网络，我们需要:

$$1 + 2 + 2 + 4 + 2 = 11$$

4.为了得到网络中变量集合的联合概率我们应用了幻灯片上的链式法则。我们有:

$$P(\neg m, \neg t, \neg h, \neg s, \neg c) = P(\neg h|\neg t) \cdot P(\neg c|\neg s, \neg t) \cdot P(\neg t|\neg m) \cdot P(\neg s|\neg m) \cdot P(\neg m)$$

= 0.9 • 0.8 • 0.9 • 0.99 • 0.3

$$= 0.192456$$

注意，上面第二行中的 0.15 是从 1−P(c|s, t)计算出来的。类似地，0.3 从 1−P(t|m)计算。

5.从幻灯片中我们可以看到:

$$\begin{aligned}\mathbf{P}(M|h,s) &= \frac{\mathbf{P}(M,h,s)}{P(h,s)} \\ &= \alpha\mathbf{P}(M,h,s) \\ &= \alpha\sum_t\sum_c \mathbf{P}(M,h,s,t,c)\end{aligned}$$

在第 4 季度，因式分解后，我们得到:

$$\mathbf{P}(M|h,s) = \alpha \sum_t \sum_c P(h|t) \cdot P(c|s,t) \cdot \mathbf{P}(t|M) \odot \mathbf{P}(s|M) \odot \mathbf{P}(M)$$
$$= \alpha \mathbf{P}(M) \odot \mathbf{P}(s|M) \odot \sum_t \sum_c P(h|t) \cdot P(c|s,t) \cdot \mathbf{P}(t|M)$$
$$= \alpha \begin{matrix} ? \\ P(\neg m) \cdot P(s|\neg m) \end{matrix} \begin{matrix} \text{P (m)} \cdot \text{P (s |米)} \sum_t \sum_c \text{P(h|t)} \cdot \text{P(c|s, t)} \cdot \text{P(t|m)} \text{ P P ?} \\ \sum_t \sum_c P(h|t) \cdot P(c|s,t) \cdot P(t|\neg m) \end{matrix}$$

假设:

$$pm = P(m) \cdot P(s|m) \sum_t \sum_c P(h|t) \cdot P(c|s,t) \cdot P(t|m)$$
$$= P(m) \cdot P(s|m) \, (P(h|t) \cdot P(t|m) \cdot P(c|s,t) + P(h|t) \cdot P(t|m) \cdot P(\neg c|s,t) +$$
$$P(h|\neg t) \cdot P(\neg t|m) \cdot P(c|s,\neg t) + P(h|\neg t) \cdot P(\neg t|m) \cdot P(\neg c|s,\neg t))$$

= 0.1 * 0.8 (0.9 * 0.7 * 0.95 + 0.9 * 0.7 * 0.05 + 0.7 * 0.3 * 0.85 + 0.7 * 0.3 * 0.15) = 0.0672

同样,我们:

$$pm' = P(\neg m) \cdot P(s|\neg m) \sum_t \sum_c P(h|t) \cdot P(c|s,t) \cdot P(t|\neg m)$$

= 0.9 * 0.2 (0.9 * 0.1 * 0.95 + 0.9 * 0.1 * 0.05 + 0.7 * 0.9 * 0.85 + 0.7 * 0.9 * 0.15)

$$= 0.1296$$

现在:

$$\mathbf{P}(M|h,s) = \alpha \begin{pmatrix} pm \\ pm' \end{pmatrix} = \alpha \begin{pmatrix} 0.0672 \\ 0.1296 \end{pmatrix} = \begin{pmatrix} 0.34 \\ 0.66 \end{pmatrix}$$

6.要计算第一个样本:

- 样本。

  P(m)是 0.1，所以我们要在 10 次中生成 m 一次，10 次中生成 9 次。如果我们在 0 和 1 之间选取一个概率相等的随机数，那么这个数将小于或等于 0.1 乘以 10，大于 0.1 乘以 10。

  因此，我们选择 m 是否为真或假的方法是将一个随机数(选取在 0 和 1 之间，包括 1，在这个范围内的每个数字的概率都相等)与 P(m)进行比较。如果随机数小于等于 P(m)，则 m 为真。否则 m 为假。

  我们的第一个随机数是 0.14，所以 m 是假的。

- 样本。

  给定，我们现在对给定的 s 进行抽样。P (s |¬m)是 0.2。我们的随机数是 0.57。所以 s 是假的。

- 样本 t。

  类似于前面的例子，P(t| - m) = 0.1，我们的随机数是 0.01，小于 0.1，所以 t 是正确的。

- 样本 c。

  我们对给定的 c 和 t 进行抽样。P(c| s, t) = 0.85，下一个随机数是 0.43，所以 c 是正确的。

- 样本 h。

  因为 t 是真的，所以我们需要在给定 t 的情况下抽样。P(h|t) = 0.9，我们的随机数是 0.59，所以 h 是真的。

我们有一个样本，s t h c ?.

按照同样的程序，我们依次得到了样本:

$$\neg m, s, \neg t, \neg c, h)$$
$$\neg m, \neg s, \neg t, \neg c, \neg h)$$
$$\neg m, \neg s, \neg t, \neg c, h,)$$
$$\neg m, \neg s, \neg t, \neg c, \neg h)$$

因此我们估计 P(m，ˆt, h, s, c) = 2/5，其中我们使用 Pˆ表示它是一个估计。

概率只有在多次迭代之后才会变得精确(当迭代的次数接近无穷大时，概率接近正确的值)。

7.因为我们在计算 P(m|h, s)，我们将只使用 h 和 s 为真的样本。使用拒绝抽样，我们继续之前。我们只需要考虑 h 和 s 都成立的事件。只有一个。在这个例子中，m 不成立，所以我们估计 P(m|h，ˆs) = 0。

同样，这是近似的，随着样本的增加会有所改善。事实上，如果你完成了问题要求的 5 个样本，你可能会得到一个更好的近似。

8.对于似然抽样，我们首先选择一个顺序来计算变量。我们将使用与之前相同的顺序。

然后，从随机数列表的开头开始

- 米样品

  m 是假的。(我们使用 0.14 来得到这个结果——我们将再次使用随机数重新开始。)

  根据定义，s 是正确的。

  w 设为 P(s| ∼ m)的值，则 w = 0.2。

- 样本 t

  t 是正确的。(我们用 0.01 来得到它。)

- h 根据定义为真。

  因此，我们用 P(h|t) = 0.9 更新 w。W = 0.2 • 0.9 = 0.18。

- c 的样本。

  c 是正确的。

我们有样本，s, t, h, c，权重是 0.18。

第二种样本:emm, s, t, tc, h ?，重量为 0.2 • 0.7 = 0.14。第三个样本:emm, s, t, c, h ?，重量为 0.2 • 0.7 = 0.14。第四种样本:emm, s, t, tc, h ?，重量为 0.2 • 0.7 = 0.14。第五种样品:emm, s, t, tc, h ?，重量为 0.2 • 0.7 = 0.14。所以:

$$\hat{\mathbf{P}}(T|h,s) = \alpha \begin{pmatrix} 0.18 \\ 4 \cdot 0.14 \end{pmatrix} \approx \begin{pmatrix} 0.243 \\ 0.757 \end{pmatrix}$$

我再次强调，这些值在这么少的样本下是非常接近的。经过几千个样本后，这些值相差很大。

9.同样，我不打算发布教程中这个可选部分的解决方案，但如果你做了，你可以根据问题 5 中的 P(m|h, s)的值来检查你的解决方案的正确性。