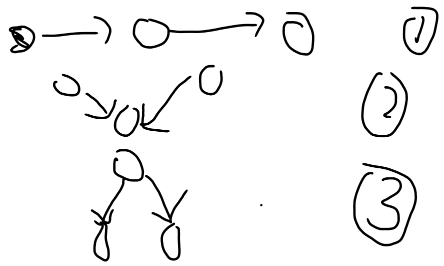
人工智能原理2

2024年6月19日 10:45

----------不确定性推理----------

1. 独立性、条件独立性



[Step 1]. Draw the ancestral graph.

根据原始概率图，构建包括表达式中包含的变量以及这些变量的ancestor节点（父节点、父节点的父节点...）的图。

[Step 2]. “Moralize” the ancestral graph by “marrying” the parents.

连接图中每个collider结构中的父节点，即若两个节点有同一个子节点，则连接这两个节点。（若一个变量的节点有多个父节点，则分别链接每一对父节点）。

[Step 3]. "Disorient" the graph by replacing the directed edges (arrows) with undirected edges (lines).

去掉图中所有的路径方向，将directional graph变为non-directional graph。

[Step 4]. Delete the givens and their edges.

从图中删除需要判断的概率表达式中作为条件的变量，以及和他们相连的路径。比如“是否P(A|BDF) = P(A|DF)?”，我们删掉D, F变量以及他们的路径。

[Step 5]. Read the answer off the graph.

- 如果变量之间没有连接，则它们在给定条件下是独立的；
- 如果变量之间有路径连接，则它们不能保证是独立的（或者粗略地说他们是不独立的，基于概率图来说）；
- 如果其中一个变量或者两者都不包含在现在的图中（作为观测条件，在step 4 被删掉了），那么他们是独立的。

例子

D-separation（D 分割）

看了三个例子，我们希望确定，任意给一个有向图即贝叶斯网络我们可以得到图中任意 A.B 是否关于 C 条件独立。

定理：在一个一般的贝叶斯网络中，A,B,C 是任意的不相交结点集合，一条路，从集合 A 中的任意的结点到集合 B 中的任意结点,如果出现下面的条件，则称为被堵住（be Blocked)

1）箭头在这条路上遇见了数据集合 C 的结点，并且这个结点是尾巴对尾巴（tail-to-tail)类型或者头对尾巴类型（head-to-tail)
2）箭头遇见了一个头对头(head-to-head)结点，这个结点既不属于集合 C,这个结点的后代也不属于集合 C

BeHappy 周琦 http://www.cnblogs.com/Dzhouqi/
如果所有从 A,到 B 的路都被堵住（Blocked),那么就称 A 与 B 被 C，D 分割了。

也可以表示为 $A \perp\!\!\!\perp B \mid C$。下图是关于后代，父代的解释。

2. 朴素贝叶斯模型：

朴素贝叶斯分类器

分类任务：N种可能的类别标记 $\mathcal{Y} = \{c_1, c_2, \ldots, c_N\}$; 要把样本 $\boldsymbol{x}$ 分类为其中一个
即要求：使得后验概率 $P(c \mid \boldsymbol{x})$ 最大的类 c

- $P(c \mid \boldsymbol{x}) = \dfrac{P(\boldsymbol{x}, c)}{P(\boldsymbol{x})} = \dfrac{P(c)P(\boldsymbol{x}|c)}{P(\boldsymbol{x})}$
- 假设样本集为 D, D_c为样本集中第c类样本组成的集合,
  - $P(c)$ 为类别的先验概率
    - 样本空间中各类样本所占比例 $P(c) = \frac{|D_c|}{|D|}$
  - $P(\boldsymbol{x}|c)$ 为样本 $\boldsymbol{x}$ 相对于类标记 c 的类条件概率，或称似然
    - 对已知类别，假设所有属性相互独立。则：
      $P(\boldsymbol{x}|c) = \prod_{i=1}^{d} P(x_i|c)$ 其中 d 为属性数目， x_i为$\boldsymbol{x}$在第i个属性上的取值
      $P(x_i|c) = \frac{|D_{c,x_i}|}{|D_c|}$ 其中 D_{c,x_i} 为D_c中在第 i 个属性上取值为x_i的样本组成的集合。
  - $P(\boldsymbol{x})$ 为用于归一化的“证据”因子,
    对于给定样本 $\boldsymbol{x}$， $P(\boldsymbol{x})$ 与类标记无关

25

拉普拉斯平滑

朴素贝叶斯分类器

需注意，若某个属性值在训练集中没有与某个类同时出现过，则直接基于式(7.17)进行概率估计，再根据式(7.15)进行判别将出现问题。例如，在使用西瓜数据集 3.0 训练朴素贝叶斯分类器时，对一个“敲声=清脆”的测试例，有

$$P_{\text{清脆}|\text{是}} = P(\text{敲声} = \text{清脆} \mid \text{好瓜} = \text{是}) = \frac{0}{8} = 0\ ,$$

由于式(7.15)的连乘式计算出的概率值为零，因此，无论该样本的其他属性是什么，哪怕在其他属性上明显像好瓜，分类的结果都将是“好瓜=否”，这显然不太合理.

为了避免其他属性携带的信息被训练集中未出现的属性值“抹去”，在估计概率值时通常要进行“平滑”(smoothing)，常用“拉普拉斯修正”(Laplacian correction)．具体来说，令 N 表示训练集 D 中可能的类别数，N_i 表示第 i 个属性可能的取值数，则式(7.16)和(7.17)分别修正为

见优缺

$$\hat{P}(c) = \frac{|D_c| + 1}{|D| + N}\ ,$$

$$\hat{P}(x_i \mid c) = \frac{|D_{c,x_i}| + 1}{|D_c| + N_i}\ .$$

例如，在本节的例子中，类先验概率可估计为

$$\hat{P}(\text{好瓜} = \text{是}) = \frac{8+1}{17+2} \approx 0.474\ , \quad \hat{P}(\text{好瓜} = \text{否}) = \frac{9+1}{17+2} \approx 0.52$$

类似地，$P_{\text{青绿}|\text{是}}$ 和 $P_{\text{青绿}|\text{否}}$ 可估计为

$$\hat{P}_{\text{青绿}|\text{是}} = \hat{P}(\text{色泽} = \text{青绿} \mid \text{好瓜} = \text{是}) = \frac{3+1}{8+3} \approx 0.364\ ,$$

优缺点：受限于独立性的假设，且无法处理小数据集情况。

朴素贝叶斯分类器的优劣

- **优势：**
  - 在大型数据集上非常容易实现并获得非常好的结果。
  - 它具有很好的计算效率。与复杂算法相比，假设所有特征都是独立的朴素贝叶斯算法非常快。在某些情况下，我们会偏好速度甚于精度。
  - 它还可以用于预测多个类别。
  - 它还可以用于自然语言处理中的文本分类。
- **劣势：**
  - 关于特征独立的强假设，这在现实生活的应用程序中几乎不成立。
  - 在数据集较小时，它的精度值会下降。
  - 当目标是预测概率而不是分类时，该方法提供非常有偏见的结果。

3. 贝叶斯网络

（1）基本术语：BN，CPT

（2）1. 贝叶斯网络的本质：

是一种数据结构，用来表示变量之间的依赖关系，可以表示任何完全联合概率分布。

2. 组成：

- 节点：对应一个随机变量，也会包含一个条件概率表，条件概率表中的每一行可能对应一个事件，那我们想要某个事件的概率，则可利用条件概率表来求
- 边：表示依赖关系

3. 构造贝叶斯网络：
   - 将变量排序，使其有一定拓扑结构，尽量保持因果顺序
   - 选择父节点，插入边
   - 记录条件概率表

4. 条件独立性：
   - 给定父节点，其独立于所有非子孙节点
   - 给定马尔科夫覆盖（父节点，子节点，子节点的父节点），独立于所有其他变量。

5. 精确推断：通过贝叶斯网络求得事件发生的概率，往往包含证据变量和查询变量，计算给定证据变量，求查询变量的后验概率分布。

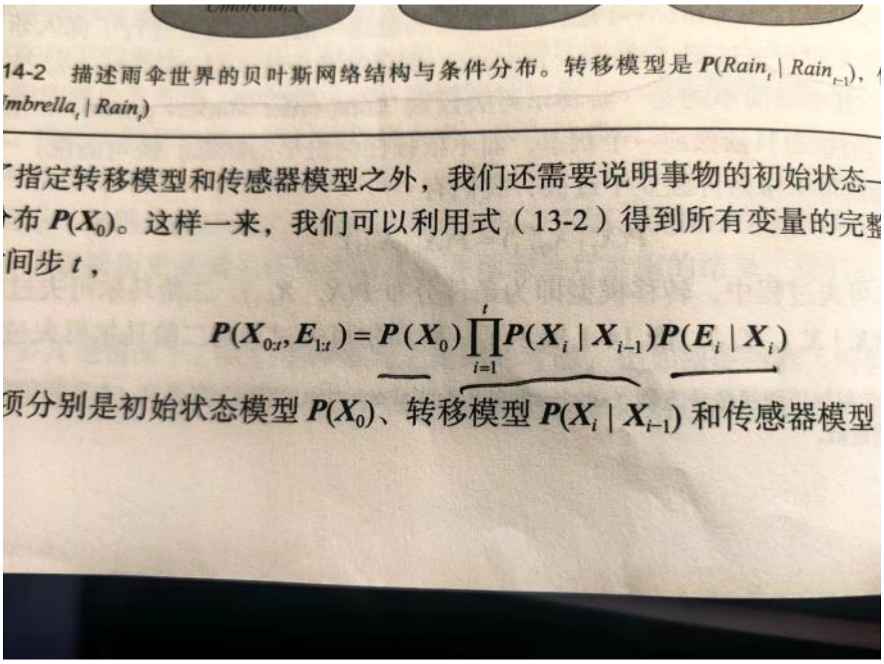
6. 如何精确推断：
   - 枚举
   - 变量消元算法
   - 将变量排序以减少复杂度
   - 聚类以减少复杂度（较为困难）

7. 近似推理：采用近似的方法求得所需事件发生概率，实际上是在使用蒙特卡洛算法进行随机采样
   - 直接采样法：按照拓扑顺序依次采样每个变量
   - 拒绝采样：先在指定先验概率中生产样本，然后删掉与证据不匹配的样本
   - 重要性采样（似然加权）：也是先按拓扑排序采样变量，当遇到证据变量时，权重乘上在给定其父变量之后的条件概率
   - 吉布斯采样：从任意一个状态出发，通过为非证据变量采样一个值来生成下一个状态。

4. 时序概率推理

（1）基本概念：

转移模型和传感器模型，一阶二阶马尔科夫过程


14-2 描述雨伞世界的贝叶斯网络结构与条件分布。转移模型是 $P(Rain_t \mid Rain_{t-1})$，

$Umbrella_t \mid Rain_t)$

了指定转移模型和传感器模型之外，我们还需要说明事物的初始状态—

布 $P(X_0)$。这样一来，我们可以利用式（13-2）得到所有变量的完整

间步 t，

$$P(X_{0:t}, E_{1:t}) = P(X_0)\prod_{i=1}^{t} P(X_i \mid X_{i-1}) P(E_i \mid X_i)$$

项分别是初始状态模型 $P(X_0)$、转移模型 $P(X_i \mid X_{i-1})$ 和传感器模型

（2）时序模型中的推断：

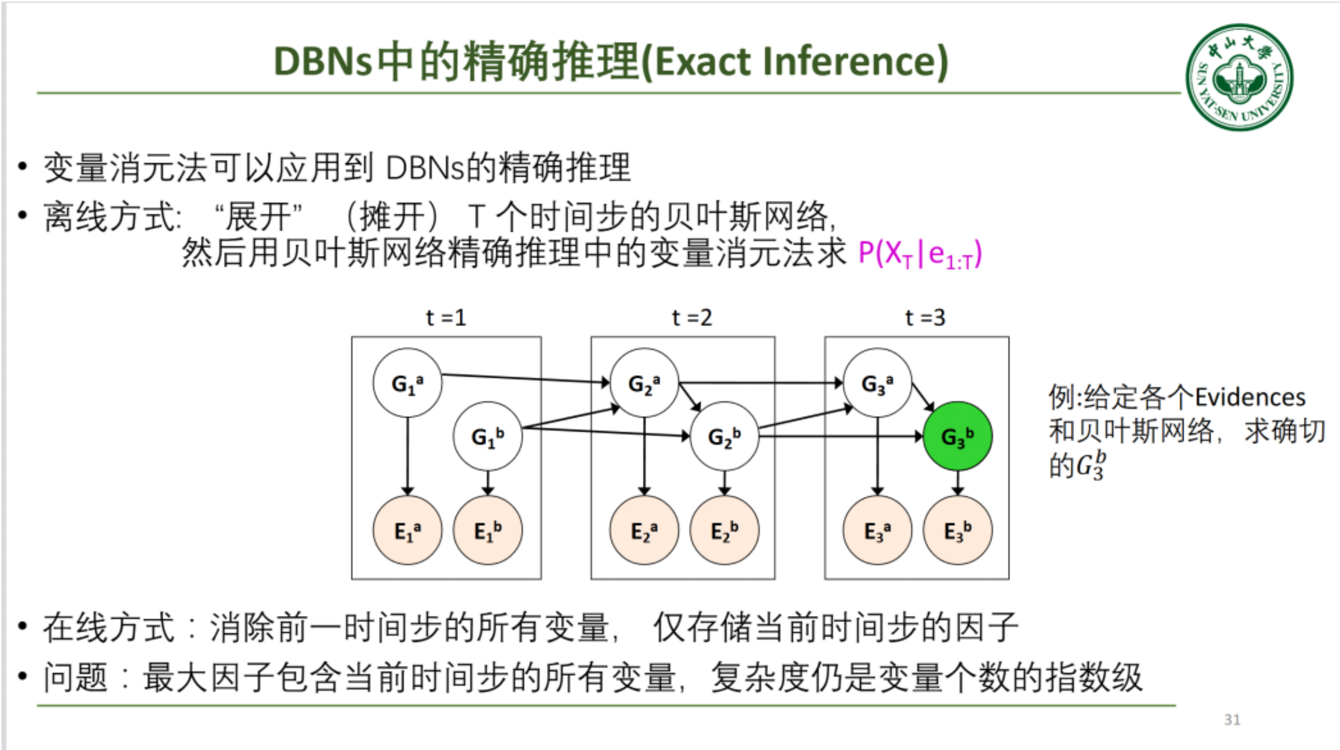
- 滤波：给定至今为止（0-t)所有证据，求t时刻变量X后验概率
- 预测：给定至今为止（0-t)所有证据，求未来K天后变量X后验概率
- 平滑：给定至今为止（0-t)所有证据，求过去第K天，K<t的变量X后验概率
- 最可能解释：给定至今为止（0-t)所有证据，找到最有可能产生这些证据的状态序列，即X_{0-t}。（有个微比特算法）
- 学习（当不知道转移模型和传感器模型时，可以从观测中学习）

（3）隐马尔科夫模型：要求过程状态由单个离散随机变量描述，对证据变量没有要求

（4）卡尔曼滤波：处理连续状态变量

（5）动态贝叶斯网络：每个时间片可以有任意数量的状态变量和证据变量，其与HMM的区别在于他可以将复杂系统的状态分解为其组成变量，以利用时序概率模型中的稀疏性。

- 精确推理：变量消元法，分为在线和离线方式

- 

- 近似推理：粒子滤波
  - 两个创新点：使用样本本身作为当前状态分布的近似表示，聚焦与状态空间的高概率区域上的样本集合
  - 步骤：预测、更新权值、重采样