不确定性推理

- 概率、独立性和贝叶斯规则复习
- 贝叶斯网络：概念、构造、推理和独立性
- 变量消除算法

*Slides based on those of Sheila McIlraith

不确定性

- 在搜索中，我们把动作看作是确定性的
  - 在状态S_1执行动作A导致变迁到状态S_2
- 而且，有一个确定的初始状态S_0。
- 所以在执行任何一系列动作之后，我们都能准确地知道我们已经到了什么状态。
- 这些假设在某些领域是合理的，但在很多领域域它们不成立。

不确定性

- 我们可能不知道开始时的状态
  - *e.g.*, 在扑克游戏中我们看不到对手的牌
  - 我们不知道病人得了什么病。
- 我们可能不知道一个动作的所有效果
  - 动作可能具有随机成分，如掷骰子。
  - 我们可能不知道一种药物的所有长期效果。
  - 动作可能会失败

不确定性

- 在这些领域，我们仍然需要采取行动，
- 但我们不能仅仅根据已知的事实采取行动。
- 我们必须“赌博”。
- 但我们如何理性地赌博呢?

一个示例

我们要去机场。但我们不确定去机场的路上交通是否拥堵。我们什么时候离开？

- 如果我们必须在工作日晚上9点到达机场。
  - 我们可以提前1小时“安全”地前往机场。
  - 出行可能需要更长的时间，但概率很低。
- 如果我们必须在星期五下午4:30到达机场。
  - 我们很可能需要1.5个小时或更长时间才能到达机场。

不确定性

- 为了在不确定性下理性地行动，我们必须能够评估某些事情的可能性。
- 通过权衡事件的可能性（概率），我们可以设计在不确定性下合理行动的机制。

概率（有限集上的）

- 概率是定义在一个事件集U，通常被称为事件的全域，上的函数。
- 它给每个事件$e \in U$ 赋值$Pr(e) \in [0,1]$。
- 它通过对事件集合F 的成员的概率求和为该集合赋值：$Pr(F) = \sum_{e \in F} Pr(e)$。
- 因此，$Pr(U) = 1, Pr(\emptyset) = 0$
- $Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B)$

一般概率

给定一个集合U (全域), 概率函数是在U的子集上定义的函数，它将每个子集映射到实数，并满足概率公理：

- $Pr(U) = 1$
- $Pr(A) \in [0, 1]$
- $Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B)$

特征向量上的概率

- 我们将处理由特征值向量组成的全域。
- 和CSPs 一样，我们有
  - 变量集合$V_1, V_2, \ldots, V_n$
  - 每个变量的有限值域,
    Dom[V_1], Dom[V_2],. . . , Dom[V_n].
- 事件的全域U 是变量值的所有向量的集合
  $\{\langle d_1, d_2, \ldots, d_n \rangle \mid d_i \in Dom[V_i]\}$
- 事件空间大小为$\prod_i |Dom[V_i]|$, *i.e.*, 论域大小的乘积.
- *e.g.*, 如果$|Dom[V_i]| = 2$, 我们有2^n 个不同的原子事件。(指数多个!)

特征向量上的概率

- 通过指定某些变量的值可以方便地说明U的一些子集, *e.g.*
  - $\{V_1 = a\}$ 表示$V_1 = a$的所有事件的集合
  - $\{V_1 = a, V_3 = d\}$ 表示$V_1 = a$并且$V_3 = d$的所有事件的集合
- 如果我们有每个原子事件（变量的完整实例化）的概率，我们就可以计算出任何集合的概率，*e.g.*
  $Pr(\{V_1 = a\}) =$
  $\sum_{x_2 \in D[V_2]} \cdots \sum_{x_n \in D[V_n]} Pr(V_1 = a, V_2 = x_2, \ldots, V_n = x_n)$

问题和解

问题

- 需要说明指数多个原子事件的概率。
- 需要将指数多个数值相加。

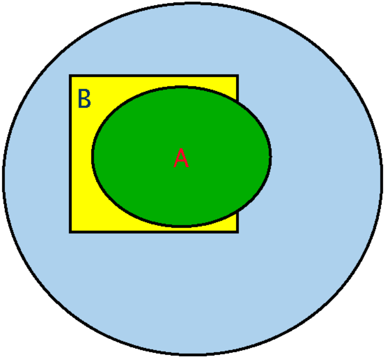
解

- 利用概率独立性，特别是条件独立性。

条件概率

- 假设A是一个事件的集合s.t. $Pr(A) > 0$.
- 那么可以定义关于事件A的条件概率:
  $Pr(B|A) = Pr(B \cap A)/Pr(A)$
- 以A为条件，对应于将注意力限制在A中的事件上。

一个示例



B覆盖了整个空间的30%左右，但覆盖了A的80%以上。
所以$Pr(B) = 0.3$, 但是$Pr(B|A) = 0.8$

性质和集合

任何事件集合A都可以解释为一个性质：具有性质A的事件集合。因此，我们经常写

- $A \vee B$来表示具有性质A或B的事件集：集合$A \cup B$
- $A \wedge B$来表示具有性质A和B的事件集：集合$A \cap B$
- $\neg A$来表示不具有性质A的事件集：集合$U - A$ (*i.e.*, A关于事件全域U的补)

求和规则

- 假设$B_1, B_2, \ldots, B_k$ 形成了全域U 的一个划分。
  - $B_i \cap B_j = \emptyset$, $i \neq j$ (互斥)
  - $B_1 \cup B_2 \cup \ldots \cup B_k = U$ (穷尽)
- 用概率表示:
  - $Pr(B_i \cap B_j) = 0$, $i \neq j$
  - $Pr(B_1 \cup B_2 \cup \ldots \cup B_k) = 1$
- 给定任意其他的事件集合A, 我们有
  $Pr(A) = Pr(A \cap B_1) + \ldots + Pr(A \cap B_k)$
- 用条件概率表示:
  $Pr(A) = Pr(A|B_1)Pr(B_1) + \ldots + Pr(A|B_k)Pr(B_k)$
- 通常我们知道$Pr(A|B_i)$, 所以我们可以这样计算$Pr(A)$。

独立性

- 有可能B在A上的密度等于它在整个集合上的密度。
  - 密度: 从整个集合中随机选择一个元素。
    所选元素在集合B 中的可能性有多大?
- 或者，B在A上的密度与它在整个空间上的密度大不相同。
- 在第一种情况下，$Pr(B|A) = Pr(B)$, 我们说B和A是独立的。
- 在这种情况下，知道一个元素属于A并不能告诉我们关于它是否也属于B的更多信息。

条件独立性

- 假设我们已经知道一个随机选择的元素具有性质A。
- 我们想知道元素是否具有性质B:
  - $Pr(B|A)$ 表示这是真的概率。
- 现在我们知道元素也有C 性质，这是否给了我们关于具有性质B的更多信息?
  - $Pr(B|A \cap C)$ 表示在附加信息下这个为真的概率。

条件独立性

- 如果$Pr(B|A \cap C) = Pr(B|A)$, 那么我们没有因为知道元素在C中而获得任何额外的信息。
- 在这种情况下，我们说给定A，B与C是条件独立的。
- 也就是说，一旦我们知道了A，额外知道C与B是否为真是无关的。
- 条件独立性是条件概率空间中的独立性$Pr(\bullet|A)$.

独立性的计算效果

- 如果A和B是独立的，则$Pr(A \cap B) = Pr(A) \cdot Pr(B)$
- 如果给定A, B和C是条件独立的，
  则$Pr(B \cap C|A) = Pr(B|A) \cdot Pr(C|A)$

Bayes规则

- Bayes规则是一个简单的数学事实。但它对于如何计算概率有很大的意义。
- $Pr(Y|X) = Pr(X|Y)Pr(Y)/Pr(X)$
- *e.g.*, 通过对心脏病患者的治疗我们可以估计出 $Pr(high\_Cholesterol|heart\_disease)$
- 使用Bayes规则，我们可以把它变成心脏病的预测指标 $Pr(heart\_disease|high\_Cholesterol)$
- 通过一个简单的血液测试，我们可以确定“高胆固醇”，并利用它来帮助估计心脏病的可能性。

Bayes 规则示例

- 疾病$\in \{malaria, cold, flu\}$; 症状= fever
- 必须计算$Pr(Disease|fever)$ 来开治疗处方
- 为什么不直接评估这个数量呢?
  - $Pr(mal|fever)$ – 评估是不自然的。它没有反映出潜在的"因果机制"疟疾$\Rightarrow$ 发热
  - $Pr(mal|fever)$ – 不"稳定": 疟疾流行会改变这个数量(例如)
- 所以我们使用Bayes 规则:
  $Pr(mal|fever) = Pr(fever|mal)Pr(mal)/Pr(fever)$

Bayes 规则示例

- 如何计算$Pr(fever)$呢？
- 假设疟疾、感冒和流感是导致发烧的唯一可能原因，*i.e.*, $Pr(fever|\neg malaria \wedge \neg cold \wedge \neg flu) = 0$, 并且它们是互斥的。
- 则$Pr(fever) = Pr(malaria \wedge fever) + Pr(cold \wedge fever) + Pr(flu \wedge fever)$
- $Pr(malaria \wedge fever) = Pr(fever|mal)Pr(mal)$
- 类似地，我们可以计算$Pr(cold \wedge fever)$ 和$Pr(flu \wedge fever)$

链式(Chain) 规则

$$Pr(A_1 \cap A_2 \cap \ldots \cap A_n) = Pr(A_1|A_2 \cap \ldots \cap A_n) \cdot$$
$$Pr(A_2|A_3 \cap \ldots \cap A_n) \cdot \ldots \cdot Pr(A_{n-1}|A_n) \cdot Pr(A_n)$$

有用的公式

- 条件概率: $Pr(B|A) = Pr(B \cap A)/Pr(A)$
- 求和规则:
  假设$B_1, B_2, \ldots, B_k$ 形成U的划分。则
  $Pr(A) = Pr(A \cap B_1) + \ldots + Pr(A \cap B_k)$
- 如果A 和B 是独立的，则$Pr(A \cap B) = Pr(A) \cdot Pr(B)$
- 如果给定A, B 和C 是条件独立的，
  则$Pr(B \cap C|A) = Pr(B|A) \cdot Pr(C|A)$
- Bayes 规则: $Pr(Y|X) = Pr(X|Y)Pr(Y)/Pr(X)$
- 链式规则: $Pr(A_1 \cap A_2 \cap \ldots \cap A_n) = Pr(A_1|A_2 \cap \ldots \cap A_n) \cdot Pr(A_2|A_3 \cap \ldots \cap A_n) \cdot \ldots \cdot Pr(A_{n-1}|A_n) \cdot Pr(A_n)$

变量独立

给定变量Z，两个变量X和Y是条件独立的，
如果对所有的$x \in \mathsf{Dom}(X)$, $y \in \mathsf{Dom}(Y)$, $z \in \mathsf{Dom}(Z)$,
给定$Z = z$，$X = x$ 和$Y = y$是条件独立的，
i.e., $Pr(X = x \wedge Y = y | Z = z) =$

$$Pr(X = x | Z = z) \cdot Pr(Y = y | Z = z)$$

符号/术语

- 对变量X，$Pr(X)$ 是指X 上的（边缘）分布。
- 它对所有$d \in Dom[X]$ 指定$Pr(X = d)$
- 注意$\sum_{d \in Dom[X]} Pr(X = d) = 1$
- 且$Pr(X = d_1 \wedge X = d_2) = 0$, 对所有$d_1, d_2 \in Dom[X]$ s.t. $d_1 \neq d_2$

符号/术语

- $Pr(X|Y)$ 指X 上的条件分布族，对每个$y \in Dom(Y)$ 有一个条件分布.
- 对每个$d \in Dom[Y]$, $Pr(X|Y=d)$ 说明X 值上的一个分布:
  $Pr(X=d_1|Y=d)$,
  $Pr(X=d_2|Y=d), \ldots, Pr(X=d_n|Y=d)$,
  其中$Dom[X] = \{d_1, d_2, \ldots, d_n\}$.
- 区分$Pr(X)$ – 是一个分布和$Pr(X=d)$ ($d \in Dom[X]$) – 是一个数。
- 把$Pr(X)$看作是一个函数，它接受任何$x \in Dom[X]$ 作为参数，并返回$Pr(X=x)$。
- 类似地，把$Pr(X|Y)$看作是一个函数，它接受任何$y \in Dom[Y]$ 并返回一个分布$Pr(X|Y=y)$.

独立性的价值

- 假设布尔变量$X_1, X_2, \ldots, X_n$ 相互独立(*i.e.*, 每个子集都变量独立于任何其他子集)
- 我们可以使用n 个参数(线性)而不是$2^n - 1$ 个参数(指数)来说明完全的联合分布(所有值向量的概率函数)
- 简单地说明$Pr(X_1 = true), \ldots, Pr(X_n = true)$ (*i.e.*, $Pr(X_i = true)$ 对所有i)
- 我们可以很容易地恢复任何原始事件的概率，*e.g.*
- $Pr(X_1 \neg X_2 X_3 X_4) = Pr(X_1)(1 - Pr(X_2))Pr(X_3)Pr(X_4)$

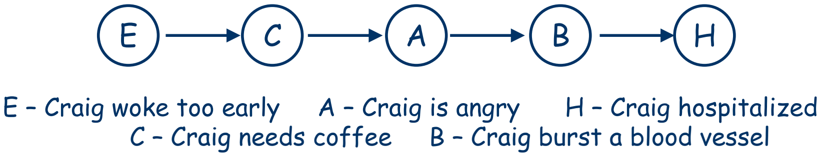
独立性的价值

- 完全独立将表示和推理从$O(2^n)$ 减少到$O(n)$!
- 然而，这种完全的相互独立是很少见的。
- 大多数现实的论域没有这种性质。
- 不过，大多数论域确实表现出相当多的条件独立性。
- 我们也可以利用条件独立性来降低表示和推断的复杂性。
- 贝叶斯网络就是用于这个目的。

利用条件独立性

考虑一个故事:

- 如果Craig 起得太早E，Craig 可能需要咖啡C;
- 如果C，他可能会生气。
- 如果A，血管破裂的几率会增加B。
- 如果B，Craig 很有可能住院H。



利用条件独立性

- 如果你知道了E、C、A或B中的任何一个，你对$Pr(H)$ 的评估就会改变。
  - e.g., 如果其中任何一个为真，将增加$Pr(h)$ 并减少$Pr(\neg h)$。
  - 所以H不独立于E，C，A，或B。
- 但如果你知道B的值(真或假)，知道了E、C或A不会影响$Pr(H)$。这些因素对H的影响是由它们对B的影响传递的。
  - Craig 不会因为生气而被送去医院，他被送去是因为他血管破裂了。
  - 因而给定B，H独立于E，C和A。

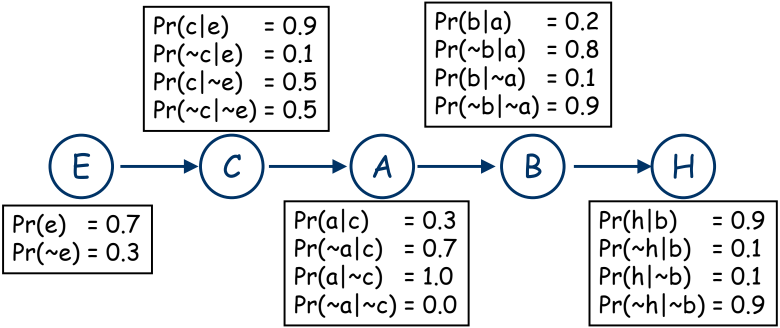
利用条件独立性

- 类似地
  - 给定A，B独立于E和C
  - 给定C，A独立于E
- 这意味着:
  - $Pr(H|B,\{A,C,E\}) = Pr(H|B)$
  - $Pr(B|A,\{C,E\}) = Pr(B|A)$
  - $Pr(A|C,\{E\}) = Pr(A|C)$
  - $Pr(C|E)$ 和$Pr(E)$ 不能简化。

利用条件独立性

- 根据链式规则(对于H，...,E的任何实例化):
  $Pr(H, B, A, C, E) =$
  $Pr(H|B, A, C, E)Pr(B|A, C, E)Pr(A|C, E)Pr(C|E)Pr(E)$
- 根据我们的独立性假设:
  $Pr(H, B, A, C, E) =$
  $Pr(H|B)Pr(B|A)Pr(A|C)Pr(C|E)Pr(E)$
- 因此，我们可以通过说明五个局部的条件分布来说明完全的联合分布:
  $Pr(H|B); Pr(B|A); Pr(A|C); Pr(C|E);$ and $Pr(E)$

示例量化



- 注意，其中一半的数值是“1减去”其他数值
- 因此，说明完全的联合分布只需要9个参数，而不是显式表示的31个参数
- 与变量的数量成线性关系，而不是指数关系！
- 如果依赖关系具有链式结构，则一般为线性

推理是容易的

想知道$P(a)$? 使用求和规则：

$$\boxed{\begin{aligned}\Pr(a) &= \sum_{c_i \in Dom(C)} \Pr(a \mid c_i)\Pr(c_i) \\ &= \sum_{c_i \in Dom(C)} \Pr(a \mid c_i) \sum_{e_i \in Dom(E)} \Pr(c_i \mid e_i)\Pr(e_i)\end{aligned}}$$

These are all terms specified in our local distributions!

推断是容易的

Pr(a)的具体计算:

- Pr(c) = Pr(c|e)Pr(e) + Pr(c|~e)Pr(~e)

  = 0.9 * 0.7 + 0.5 * 0.3 = 0.78
- Pr(~c) = Pr(~c|e)Pr(e) + Pr(~c|~e)Pr(~e) = 0.22
  - Pr(~c) = 1 – Pr(c), as well
- Pr(a) = Pr(a|c)Pr(c) + Pr(a|~c)Pr(~c)

  = 0.3 * 0.78 + 1.0 * 0.22 = 0.454
- Pr(~a) = 1 – Pr(a) = 0.546

- 上面的结构是一个贝叶斯网络。
- BN是对一组变量的直接依赖关系的图形表示，以及一组条件概率表(CPTs)，用于量化这些影响的强度。
- 贝叶斯网络推广了上述例子中的思想，从而得到不确定性下表示和推理的有效方法。

贝叶斯网络

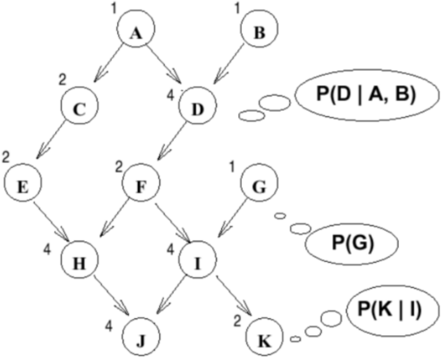
变量$\{X_1, X_2, \ldots, X_n\}$ 上的一个BN 包括:

- 一个DAG(directed acyclic graph, 有向无环图)，其节点为变量
- 一个CPTs(conditional probability tables, 条件概率表)的集合：$Pr(X_i|Par(X_i))$ 对每个X_i

关键概念

- 节点的父节点: $Par(X_i)$
- 节点的子节点
- 节点的后代
- 节点的祖先
- 家族：由X_i及其父节点组成的节点集

示例(二值变量)



- “显示”了若干CPTs
- 显式表示联合分布需要$2^{11}-1 = 2047$ 个参数
- BN只需要27个参数（列出了每个CPT的大小）

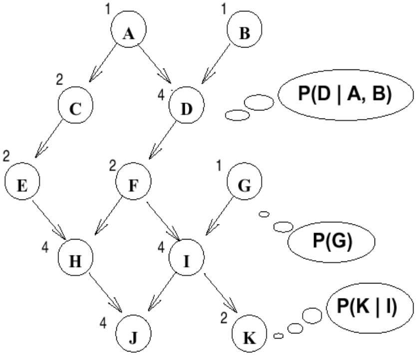
贝叶斯网络的语义

- 贝叶斯网络说明了网络中变量的联合分布可以写成下面的乘积分解。
  $Pr(X_1, X_2, \ldots, X_n) = Pr(X_n|Par(X_n)) *$
  $Pr(X_{n-1}|Par(X_{n-1})) * \ldots * Pr(X_1|Par(X_1))$
- 这个等式对于变量$X_1, X_2, \ldots, X_n$的任何取值$d_1, d_2, \ldots, d_n$都成立
- *e.g.*, 我们有X_1, X_2, X_3，每个具有论域$Dom[X_i] = \{a, b, c\}$，并且
  $Pr(X_1, X_2, X_3) = P(X_3|X_2)P(X_2)P(X_1)$
- 则$Pr(X_1 = a, X_2 = a, X_3 = a) =$
  $P(X_3 = a|X_2 = a)P(X_2 = a)P(X_1 = a)$

示例(二值变量)

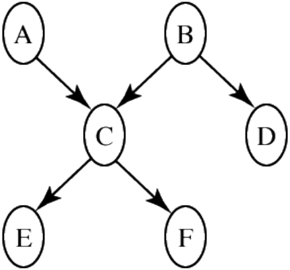
Pr(A,B,C,D,E,F,G,H,I,J,K) =

Pr(A)
x Pr(B)
x Pr(C|A)
x Pr(D|A,B)
x Pr(E|C)
x Pr(F|D)
x Pr(G)
x Pr(H|E,F)
x Pr(I|F,G)
x Pr(J|H,I)
x Pr(K|I)



练习

计算$P(c|a,b,\neg d,\neg e,\neg f)$. 注意$\neq P(c|a,b)$



$$
\begin{aligned}
P(a) &= 0.9 & P(d|b) &= 0.1\\
P(b) &= 0.2 & P(d|\neg b) &= 0.8\\
P(c|a,b) &= 0.1 & P(e|c) &= 0.7\\
P(c|a,\neg b) &= 0.8 & P(e|\neg c) &= 0.2\\
P(c|\neg a,b) &= 0.7 & P(f|c) &= 0.2\\
P(c|\neg a,\neg b) &= 0.4 & P(f|\neg c) &= 0.9
\end{aligned}
$$

给定A 和B，C 和E 不是独立的