电子科技大学
University of Electronic Science and Technology of China

5 局部概率模型
6 基于模板的表示

吴睿智

Data Mining Lab, Big Data Research Center, UESTC
Email：junmshao@uestc.edu.cn
http://staff.uestc.edu.cn/shaojunming

5 局部概率模型

本集提要

1 CPD确定性
2 特定上下文的CPD表示
3 因果影响下的独立性模型
4 连续变量和条件贝叶斯网

非特别说明，今天的讨论集中在有向图中。

CPD(条件概率分布)

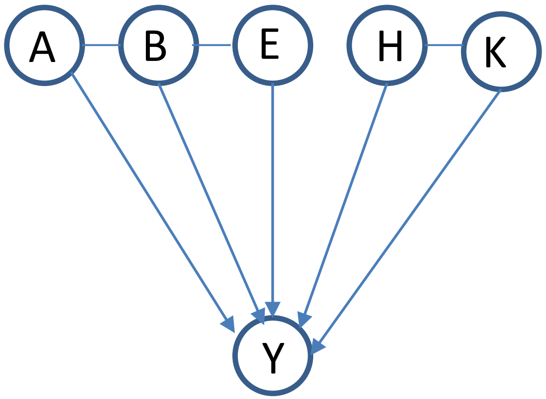
处理完全由离散值随机变量构成的空间时，总是可以借助CPD的表格表示:
假设x由a,b,c三个变量构成的父节点集合决定。

X	$a^0b^0c^0$	$a^0b^0c^1$	$a^0b^1c^0$	$a^0b^1c^1$	$a^1b^0c^0$	$a^1b^0c^1$	$a^1b^1c^0$	$a^1b^1c^1$
x^0	0.3	0.2	0.5	0.2	0.3	0.4	0.1	0.1
x^1	0.7	0.8	0.5	0.8	0.7	0.6	0.9	0.9

这个表格包含了关于x和parent(X)的每一个联合赋值的表值，为了让表格更合理，我们约束

$$\sum_{x \in Val(x)} P(x|pa_x) = 1$$

CPD(条件概率分布)



$$2^{10}=1024$$

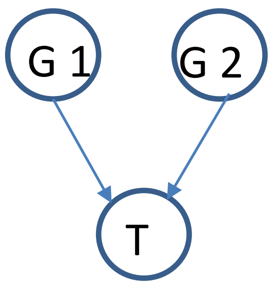
1、无法在一个表格里存储每个可能的条件概率
2、表格表示的另一个问题：忽略了CPD的内在结构。现实生活中，很多例子表明对于X的父节点的不同的值，参数之间有一定的规律性。

$$p(y\,|\,\mathrm{A})$$

第5章的大部分表示方法都试图清楚的描述并且利用一些规律性，减少具体的一个CPD的参数量

CPD中的确定性

确定性的表示



G1={a,b,o}
G2={a,b,o}

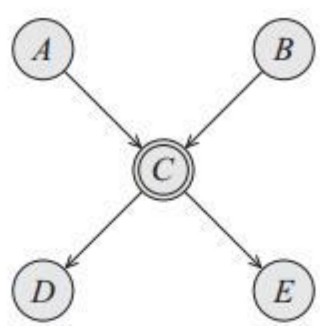
$$T=\begin{cases} ab & \text{if } G_1 \text{ or } G_2 \text{ is } a \text{ and the other is } b \\ a & \text{if at least one of } G_1 \text{ or } G_2 \text{ is equal to } a \text{ and the other is either } a \text{ or } o \\ b & \text{if at least one of } G_1 \text{ or } G_2 \text{ is equal to } b \text{ and the other is either } b \text{ or } o \\ o & \text{if } G_1 = o \text{ and } G_2 = o \end{cases}$$

确定性变量可以帮助简化复杂模型中的依赖性

独立性

条件分布概率的独立性判别是依赖数值定义的

$$P(\mathrm{x},\mathrm{y}\,|\,\mathrm{z}) = \mathrm{P}(\mathrm{x}\,|\,\mathrm{z}) \bullet P(\mathrm{y}\,|\,\mathrm{z})$$



$$A, B \models C \models (\mathrm{D} \perp \mathrm{E})$$

$$(\mathrm{D} \perp \mathrm{E} \,|\, \mathrm{A}, \mathrm{B})$$

确定性CPD

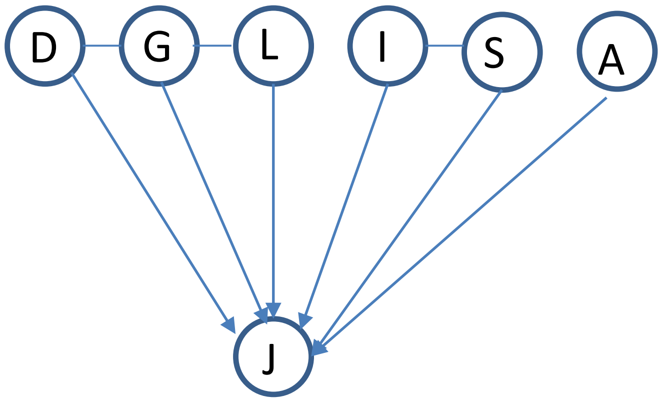
令X，Y，Z为两两互不相交的变量集合，令C是变量的一个集合(可能与X,Y,Z有交集)，且令$c \in Val(C)$

$$P(\mathrm{X} \mid \mathrm{Y}, \mathrm{Z}, \mathrm{c}) = P(\mathrm{X} \mid \mathrm{Z}, \mathrm{c})$$

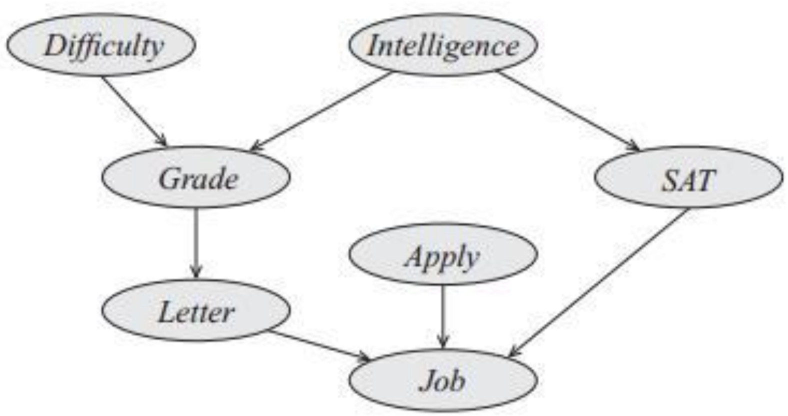
$$satisfy \quad P(\mathrm{Y}, \mathrm{Z}, \mathrm{c}) > 0$$

则称X与Y在给定Z和表示为$(\mathrm{X} \perp_c \mathrm{Y} \mid \mathrm{Z}, \mathrm{c})$的上下文c时上下文独立

CPD中的特定上下文的表示



A :Apply
S:SAT
L:Letter
G:Grade
I:Intelligence
D:Difficluty
J:Job



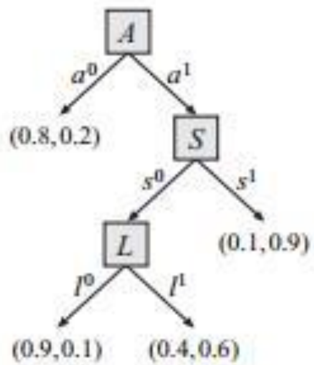
在这个有向图网络中，还是觉得变量太多，通过观察图发现对Job起作用的就是Letter、Apply、SAT这三个变量。

如果能够给一些特定的上下文，变量数能够进一步化简。这种有向图网络能够表示变量之间的依赖关系，但无法刻画具体的条件概率分布。所以引入了刻画条件分布的两种表示方法，树和规则。

树（Tree-CPDS）

Tree-CPDs

$P(J \mid A, S, L)$

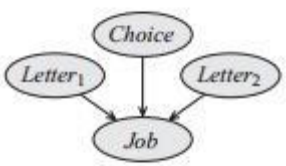


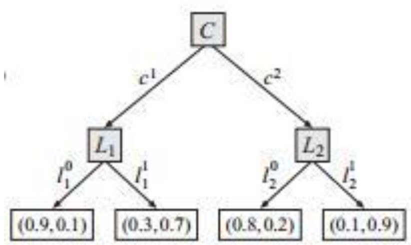
树的叶节点代表了定义在J上的不同的条件分布，通过每个叶节点的路径则决定了使用这个分布的上下文。

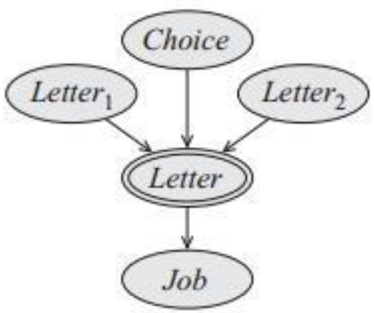
$$P(J^1 \mid a^1, s^0, l^1) = 0.6$$

$$P(J^0 \mid a^1, s^0, l^1) = 0.4$$

$P(J \mid L_1, L_2, C)$



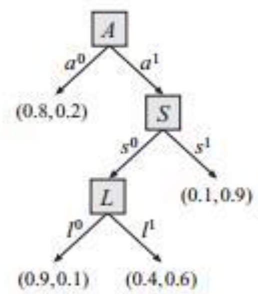




如果，$Val(A) = \{1, \ldots k\}$且 $P(Y \mid A, Z_1, \ldots Z_k) = 1\{Y = Z_a\}$其中A=a，变量A称为CPD的选择器变量，则 $CPDP(Y \mid A, Z_1, \ldots Z_k)$称为一个多路复用器CPD。

规则集(Rule—CPDS)

树结构是一种单一数据结构中描述整个CPD的全局表示。在许多情况下，如果能够把依赖结构分解为更细粒度的元素，用CPD推理会更加容易。一个上下文特定依赖性的更细粒度表示就通过规则实现。



$$\left\{\begin{array}{l}
\rho_1{:}\langle a^0, j^0; 0.8\rangle \\
\rho_2{:}\langle a^0, j^1; 0.2\rangle \\
\rho_3{:}\langle a^1, s^0, l^0, j^0; 0.9\rangle \\
\rho_4{:}\langle a^1, s^0, l^0, j^1; 0.1\rangle \\
\rho_5{:}\langle a^1, s^0, l^1, j^0; 0.4\rangle \\
\rho_6{:}\langle a^1, s^0, l^1, j^1; 0.6\rangle \\
\rho_7{:}\langle a^1, s^1, j^0; 0.1\rangle \\
\rho_8{:}\langle a^1, s^1, j^1; 0.9\rangle
\end{array}\right\}$$

定义：规则p是一个偶对<c;p>其中，c是变量C的某个子集的一个赋值，并且p在[0,1]。C定义为p的辖域，记做Scope[p]

这种表示把tree –CPD分解为其最基本的元素。

规则集(Rule—CPDS)

基于规则的CPDP(X|Parx)是满足如下条件的规则集R：

- 对每个规则p属于R， $Scope[\mathrm{p}] \in \{\mathrm{X}\} \cup Pa_x$
- 对每一个 $\{\mathrm{X}\} \cup Pa_x$ 的赋值(x,u),恰好有一个规则 $< c; p > \in R$,使得c与(x,u)匹配，在这种情况下，$P(\mathrm{X} = \mathrm{x} \mid \mathrm{Pa}_{\mathrm{x}} = \mathrm{u}) = \mathrm{p}$ 成立
- 作为结果的CPDP(X|U)是合法的CPD，因为

$$\sum_x P(\mathrm{x} \mid \mathrm{u}) = 1$$

最开始的规则集是有Tree-CPDs推导得出，那Rule_CPDS是否可以导出CPD分布呢？

X是关于Pa$_x$={A,B,C}，X的CPD根据以下规则集定义：

$\rho_1: \langle a^1, b^1, x^0; 0.1 \rangle$
$\rho_2: \langle a^1, b^1, x^1; 0.9 \rangle$
$\rho_3: \langle a^0, c^1, x^0; 0.2 \rangle$
$\rho_4: \langle a^0, c^1, x^1; 0.8 \rangle$
$\rho_5: \langle b^0, c^0, x^0; 0.3 \rangle$
$\rho_6: \langle b^0, c^0, x^1; 0.7 \rangle$
$\rho_7: \langle a^1, b^0, c^1, x^0; 0.4 \rangle$
$\rho_8: \langle a^1, b^0, c^1, x^1; 0.6 \rangle$
$\rho_9: \langle a^0, b^1, c^0; 0.5 \rangle$

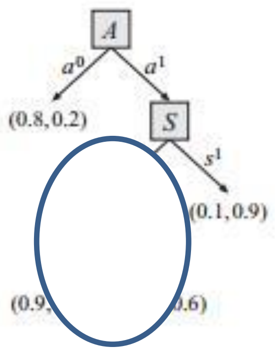
X	a1b1c0	a1b1c1
x0	0.1	0.1
x1	0.9	0.9

X	$a^0b^0c^0$	$a^0b^0c^1$	$a^0b^1c^0$	$a^0b^1c^1$	$a^1b^0c^0$	$a^1b^0c^1$	$a^1b^1c^0$	$a^1b^1c^1$
x^0	0.3	0.2	0.5	0.2	0.3	0.4	0.1	0.1
x^1	0.7	0.8	0.5	0.8	0.7	0.6	0.9	0.9

CPD表值中的每个表值都恰好与一个上下文匹配，不同上下文时相互排斥穷尽的。

确定性CPD中的独立性与约简

独立性



给定上下文S1已知，显然，可以只考虑与这个值一致的树枝，这里存在两条树枝，一条是a0，一条是a1s1，这两条树枝之间的选择并不依赖L的值，因此能够得出：

$$(J \perp_c L \mid S^1)$$

约简

$\rho_1:\langle a^1, b^1, x^0; 0.1\rangle$
$\rho_3:\langle a^0, c^1, x^0; 0.2\rangle$
$\rho_5:\langle b^0, c^0, x^0; 0.3\rangle$
$\rho_7:\langle a^1, b^0, c^1, x^0; 0.4\rangle$
$\rho_9:\langle a^0, b^1, c^0; 0.5\rangle$

$\rho_2:\langle a^1, b^1, x^1; 0.9\rangle$
$\rho_4:\langle a^0, c^1, x^1; 0.8\rangle$
$\rho_6:\langle b^0, c^0, x^1; 0.7\rangle$
$\rho_8:\langle a^1, b^0, c^1, x^1; 0.6\rangle$

$\xrightarrow{a1}$

$\rho'_1:\langle b^1, x^0; 0.1\rangle$
$\rho_5:\langle b^0, c^0, x^0; 0.3\rangle$
$\rho'_7:\langle b^0, c^1, x^0; 0.4\rangle$

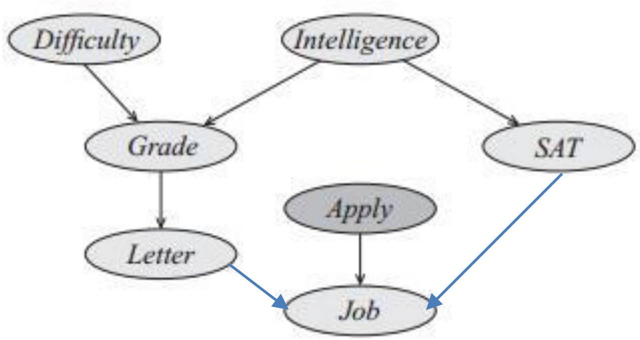
$\rho_2:\langle b^1, x^1; 0.9\rangle$
$\rho_6:\langle b^0, c^0, x^1; 0.7\rangle$
$\rho'_8:\langle b^0, c^1, x^1; 0.6\rangle$.

令 $\rho = <c';p>$是一个规则，C=c是一个上下文。如果c’与c匹配，则称 $\rho \sim c$ 。令 $c'' = c' < Scope[c'] - Scope[c] >$ 是c’中对 $Scope[c'] - Scope[c]$ 中变量的赋值。定义约简规则 $\rho[c] = <c''; p>$ 。对规则集R，约简规则集定义为

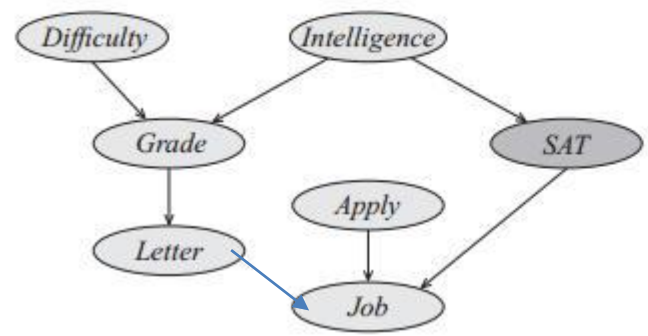
$$R[c] = \{\rho[c] : \rho \in R, \rho \sim c\}$$

确定性CPD中的去伪

令P(X|Pax)是一个CPD，$Y \in Pa_x$ 并且令c是一个上下文。如果P(X|Pax)满足$(X \perp_c Y \mid Pa_x - \{Y\}, c')$，则称边 $Y \to X$ 在上下文c中是伪的，其中 $c' = c < Pa_x >$ 是c对Pax中变量的约束。

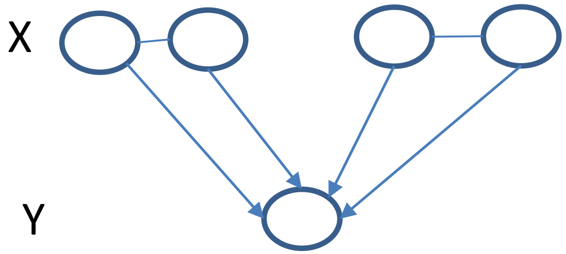


$$A = a^0$$



$$S = s^1$$

因果影响下的独立性模型



考虑局部概率中一个非常不同的结构，变量Y,分布取决于某些原因X1，X2。。。Xk，Xi之间能够以复杂的方式互相作用，从而使每种组合的影响都与其他组合无关。

有些情况，Xi关于Y的组合影响是每个Xi关于Y的组合影响是每个Xi关于Y的单独影响的简单组合。有些情况是Xi之间的交互关系，而且交互关系并不清楚，但是他们单独引起结果发生的概率是清楚的，用两个模型noisy-or模型和广义线性模型来分别刻画这两种情况。

Noisy-or

邵老师给推荐信有两个要求，1、你平常课堂表现很积极，或者，2、你最后的大论文非常赞。他就会给你一个棒棒哒的推荐信。但是这两个条件之间的交互影响在邵老师心里也非常难刻画。用一个CPD分布来描述这个事情，Q代表课堂表现，F代表大论文。

Q, F	l^0	l^1
$q^0 f^0$	1	0
$q^0 f^1$	0.1	0.9
$q^1 f^0$	0.2	0.8
$q^1 f^1$	0.02	0.98

这种原因间的交互作用称为noisy-or模型。

令Y是k个二值父节点*X*1, . . . , *Xk*.的一个二值随机变量，如果存在满足：

$$P(\mathrm{y}^0 \mid X_1,\ldots\ldots \mathrm{X}_\mathrm{k}) = (1-\lambda_0)\prod_{i:X_i=x_i^1}(1-\lambda_\mathrm{i})$$

$$P(\mathrm{y}^1 \mid X_1,\ldots\ldots \mathrm{X}_\mathrm{k}) = 1-[(1-\lambda_0)\prod_{i:X_i=x_i^1}(1-\lambda_\mathrm{i})]$$

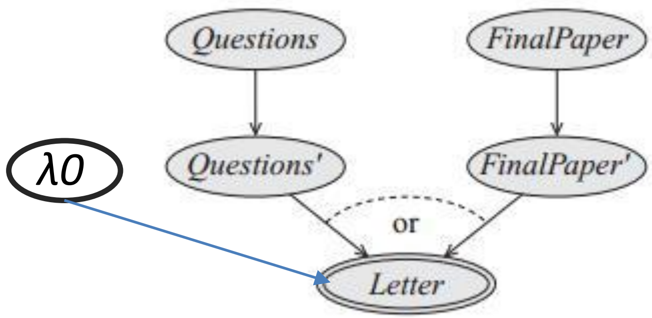
的k+1个噪声系数*λ*0, *λ*1, . . . , *λk*，则CPD *P*(*Y* | *X*1, . . . , *Xk*) 是一个noisy-or CPD

其中一个原因单独引起结果的概率称为噪声系数。

Noisy-or

Q, F	l^0	l^1	
$q^0 f^0$	1	0	
$q^0 f^1$	0.1	0.9	λ_f
$q^1 f^0$	0.2	0.8	λ_Q
$q^1 f^1$	0.02	0.98	

理解这种交互作用的另一个方法是用一个概率图模型来表示。图中会有两个Q和两个F。



泄露概率λ

泄露概率的补1-λ

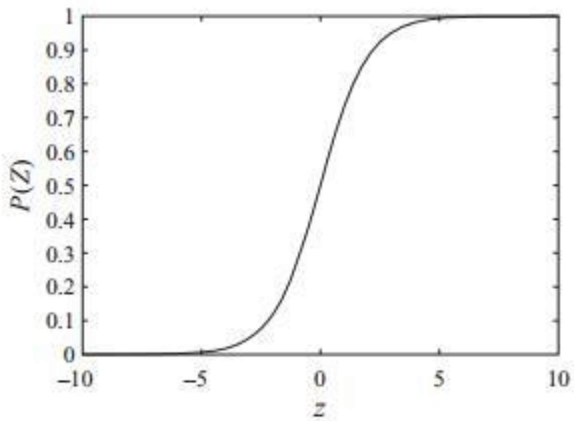
λ0表示泄露概率，比如，λ_0=0.000001 峰峰在上课没去听过课没去交过大论文，但邵老师仅仅因为今天中午的食堂菜比较合胃口就写了推荐信。这个事件可以简单引起一个变量表示，该变量没有父节点。

Generalized Linear Models

广义线性模型

在给定X1,X2,..Xk的条件下，检查Y的CPD。假定Xi对Y的影响可以通过线性函数 $f(X_1,\ldots X_k)=w_0+\sum_{i=1}^{k}w_iX_i$ 来概括，各因素之间无交互作用。当 $f(\mathrm{X}_1,\mathrm{X}_2\ldots\mathrm{X}_k)\geq\Gamma$ Y=1，当 $f(\mathrm{X}_1,\mathrm{X}_2\ldots\mathrm{X}_k)\leq\Gamma$ Y=0。

$$sigmoid(\mathrm{z})=\frac{e^z}{1+e^z}$$



令 Y 是一个有k个取数值父节点 X1, . . . , Xk 二值变量。如果存在k+1个权重w0, w1, . . . , wk 满足 $P(\mathrm{y}_1\mid X_1,\ldots X_k)=\mathrm{sigmoid}(w_0+\sum_{i=1}^{k}w_iX_i)$

CPD P(*Y* | *X*1, . . . , *Xk*) logistic CPD。

连续变量

令Y是有连续父节点$X_1, . . . , X_k$的连续变量。假如存在参数 $\beta_0, . . . , \beta_k$ 和 σ2 使得

$$p(Y \mid x_1, . . . , x_k) = N(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k; \sigma^2)$$

则称Y有一个线性高斯模型，该模型可以用向量表达：

$$p(Y \mid x) = N(\beta_0 + \beta^T x; \sigma^2)$$

公式表明，Y是变量$X_1, . . . , X_k$，外加均值为0，方差为σ2的高斯噪声的线性函数：

$$Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \varepsilon$$

混合变量

令X是一个连续变量，并且令U={U1,..Um}是它的离散父节点，Y={Y1,...Yk}是它的连续父节点。如果对任意值 $u \subseteq Val(\mathrm{U})$ ，存在包含k+1个参数au0,...auk的集合和方差 σ_u^2 ，使得

$$p(\mathrm{X} \mid \mathrm{u}, \mathrm{y}) = N(\mathrm{a}_{\mathrm{u},0} + \sum_{i=1}^{k} a_{u,i} y_i ; \sigma_u^2)$$

则称X具有条件线性高斯CPD（CLG CPD）

条件贝叶斯网

在给定X的条件下，Y上条件贝叶斯网B定义为节点是 $X \cup Y \cup Z$ 的一个有向无圈图G，其中X,Y,Z互不相交。X中的变量称为输入，Y中的变量称为输出，而Z中的变量是封装的，X中的变量在G中没有父节点。$Z \cup Y$ 中的变量与一个CPD关联。网络用链式规则定义了如下的条件分布

$$P_B(\mathrm{Y}, \mathrm{Z} \mid \mathrm{X}) = \prod_{c \in Y \cup Z} P(\mathrm{C} \mid \mathrm{Pa}_c^G)$$

分布 $P_B(\mathrm{Y} \mid \mathrm{X})$ 定义为 $P_B(\mathrm{Y}, \mathrm{Z} \mid \mathrm{X})$ 的边缘

$$P_B(\mathrm{Y} \mid \mathrm{X}) = \sum_{Z} P_B(\mathrm{Y}, \mathrm{Z} \mid \mathrm{X})$$

6 基于模板的表示

本集提醒

1 时序模型
2 对象关系领域的有向概率模型

时序模型

数据挖掘实验室
Data Mining Lab

时序模型的基本假设

$$P(\chi^{(0)},\chi^{(1)},\ldots\chi^{(\mathrm{T})})$$

在预先指定的采样时间t=0,...T的轨迹上考虑分布 $P(\chi^{(0)},\chi^{(1)},\ldots\chi^{(\mathrm{T})})$ 该分布通常简写为 $P(X^{(0:T)})$。在与时间一直的方向上，用概率的链式法则可以将分布重新参数化

$$P(\chi^{(0:\mathrm{T})})=P(\chi^{(0)})\prod_{t=0}^{T-1}p(\chi^{(\mathrm{t}+1)}\mid\chi^{(0:\mathrm{t})})$$

对模板变量X的动态系统，如果对于所有$t\geq 0$，如果下式成立：

$$(\chi^{(\mathrm{t}+1)}\perp\chi^{(0:\mathrm{t}-1)}\mid\chi^{(\mathrm{t})})$$

则称该动态系统满足马尔科夫假设，并称该系统为马尔科夫系统。
分布定义为更紧凑的表示：

$$P(\chi^{(0)},\chi^{(1)},\quad\chi^{(\mathrm{t})})=\mathrm{P}(\mathrm{x}^{(0)})\prod_{t=0}^{T-1}P(\chi^{(\mathrm{t}+1)}\mid\chi^{(\mathrm{t})})$$

数据挖掘实验室

Data Mining Lab

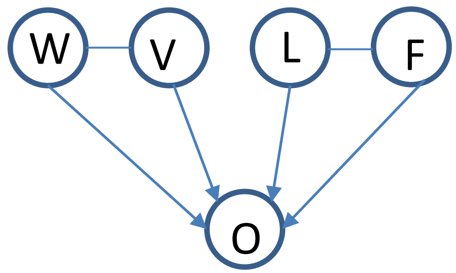
平稳马尔科夫系统

如果对所有的t，$P(X^{(t+1)} | X^{(t)})$相同，那么马尔科夫动态系统称为平稳的。在这种情况下，该过程可以用转移模型表示 $P(X' | X)$，并且使得对于任意的 $t \geq 0$

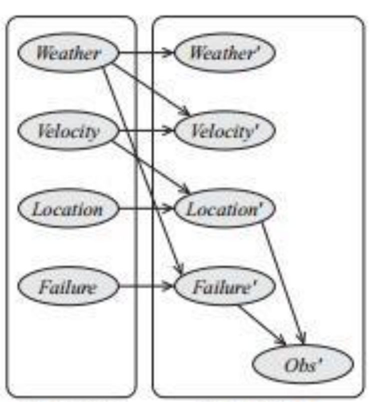
$$P(\chi^{(t+1)} = \varepsilon' | \chi^{(t)} = \varepsilon) = P(\chi' = \varepsilon' | \chi = \varepsilon)$$

动态贝叶斯网络

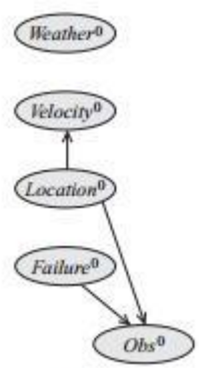
汽车的位置依赖于天气 速度 先前的位置 以及汽车是否有故障等条件，下一时刻的位置只与上一时刻的汽车状态和位置有关。但通常情况下，我们只能观察到当前汽车所处的位置就是观测值，以上四个原因的内部关系是未知的。



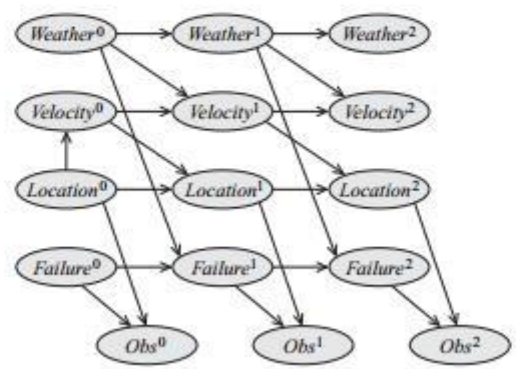
所以要构建一个能够描述汽车位置的网络来刻画这个模型。



Time slice t Time slice $t+1$



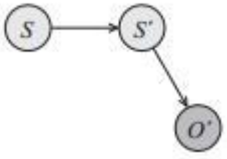
Time slice 0



Time slice 0 Time slice 1 Time slice 2

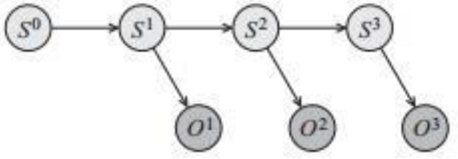
动态贝叶斯网络

对于集合χ上的一个过程，在给定χ_I的条件下，χ'上的条件贝叶斯网是一个2-时间片段贝叶斯网（2-TBN），其中$\chi_I \subseteq \chi$是界面变量的集合。



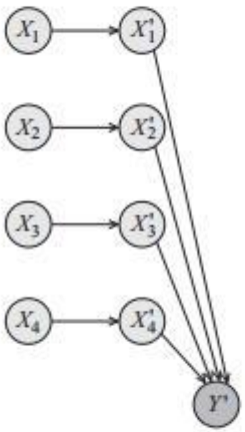
动态贝叶斯网是一个偶对$<B_0, B_\rightarrow>$，其中B_0是$\chi^{(0)}$上表示状态的初始分布的贝叶斯网，$B_\rightarrow$是过程的一个2-TBN。对任意需要的时间跨度$T \geq 0$，$\chi^{(0:T)}$上的分布定义为展开的贝叶斯网，其中对任意的i=1,……n：

- $X_i^{(0)}$的结构和CPD与X_i在B_0一致。
- 对t>0，X_i^t的结构和CPD与X_i'在$B_\rightarrow$中的一致。

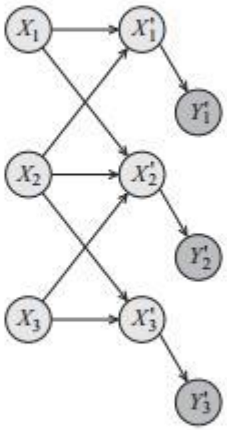


因子HMM，它的2-TBN具有一系列链式结构 $X_i \rightarrow X_i^{'}$(i=1,....N),
通常具有一个单一的观测变量Y’，它是所有变量 $X_i^{'}$ 的子节点。

耦合的HMM同一由一系列X_i链构造，不过现在每条链是有其自己私有观测变量Y_i的一个HMM。这些链通过状态变量直接交互作用，每条链都对与其相邻的链产生影响。



因子HMM



耦合HMM

状态观测模型

另一种考虑时序过程的方法是状态-观测模型(State-Observation Models).在状态观测模型中，就其在单独过程中发生的观测，系统可以视为其自身的自然进化。

状态变量以马尔科夫方式进化，所以 $(X^{(t+1)} \perp X^{(0:(t-1))} \mid X^{(t)})$

给定t时刻的状态时，t时刻的观测变量与整个状态序列条件独立：

$$(O^{(t)} \perp X^{(0:(t-1))}, X^{(t+1:\infty)} \mid X^{(t)})$$

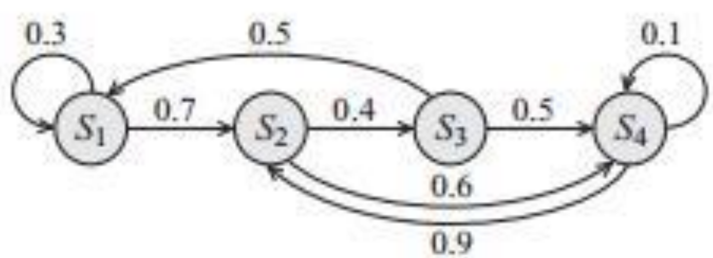
这种概率模型由两部分组成：组成模型P(X’|X)和转移模型P(O|X)

状态观测模型

考虑一个HMM，其唯一的状态S可取四个值s1,s2,s3,s4,其转移模型：

	s_1	s_2	s_3	s_4
s_1	0.3	0.7	0	0
s_2	0	0	0.4	0.6
s_3	0.5	0	0	0.5
s_4	0	0.9	0	0.1

HMM转移模型用有向图（通常有环）来进行编码表示，图的节点代表系统的不同状态，也就是Val(S)中的值。如果有可能从s转移到s’，P(s|s’)>0,那么从s到s’就有一条弧。还可以将从s到s’的边标注为与其相关的转移概率P(s|s’)。



状态观测模型

线性动态系统

$$P(X^{(t)} \mid X^{(t-1)}) = N(AX^{(t-1)}; Q) \qquad P(O^{(t)} \mid X^{(t)}) = N(H\,X^{(t)}; R)$$

P不是概率，是分布！！！

X是状态变量的n维向量，O是观测变量的m维向量，A是用于定义线性模型的n*n矩阵，Q是用于定义与系统相关的高斯噪声的n*n矩阵，H是用于定义线性观测模型的n*m矩阵，R是个用于定义与观测相关的高斯噪声的m*m矩阵。

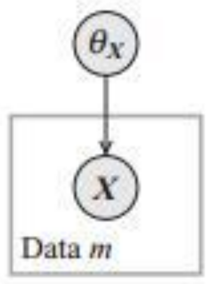
上面的状态和观测变量仍然是实数的一个向量，但是转移和观测模型是非线性函数而不是上面的线性矩阵乘法时，就有：

$$P(X^{(t)} \mid X^{(t-1)}) = f(X^{(t-1)}, U^{(t-1)}) \qquad P(O^{(t)} \mid X^{(t)}) = g(X^{(t)}, W^{(t)})$$

f和g是确定的非线性函数。

Plate 模型

X是投掷一枚硬币（可能有偏置）的结果，θ_x 在区间[0,1]上取值，并且其值代表了硬币的偏置程度（正面出现的概率）。



描述了从同一个分布中生成的多个随机变量。存在一个所有变量都从同一个分布中采样并且具有相同值域Val(X)的随机变量集X(d) ($d \in D$)。在plate表示中，通过只画出一个单一阶段X(d)，并将其封闭在一个方框中可以表示d包含于 D，表达了这些变量全部产生于同一个模板的的事实，该方框称为一个plate。

将参数直接包含在模型中，可以使所有变量X(d)都从同一个CPD中采样这一事实更加明确。

Plate 模型

对每个具有参数名$U_1,\ldots U_k$的模板属性 $A \in \aleph$，plate模型M_{plate}定义了：

- 一个模板父节点的集合

$$Pa_A = \{\mathrm{B}_1(\mathrm{U}_1), \ldots \mathrm{B}_\mathrm{L}(\mathrm{U}_\mathrm{L})\}$$

使得对每个$B_i(U_i)$,有 $U_i \subseteq \{\mathrm{U}_1, \ldots \mathrm{U}_\mathrm{k}\}$ 。变量U_i是父节点B_i的参数名。

- 一个模板CPD$P(A|Pa_A)$。

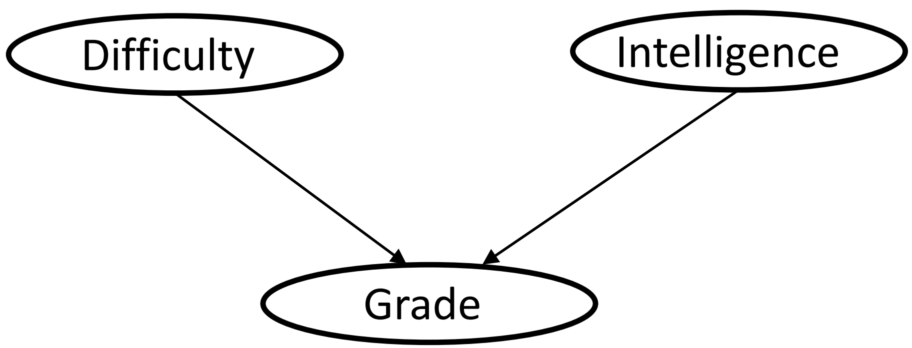
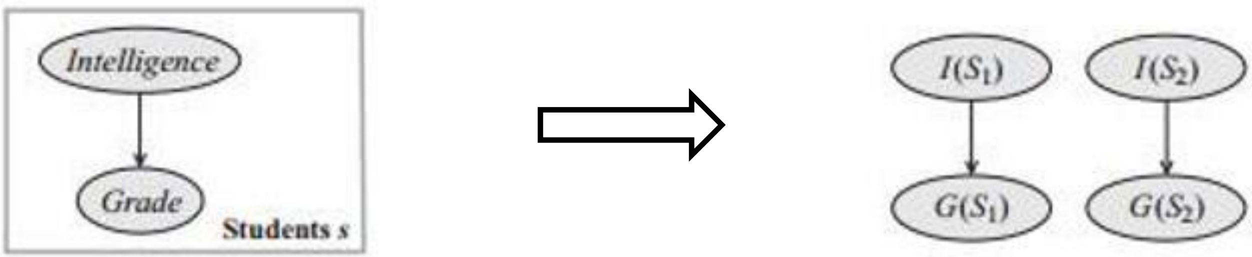


Plate 模型和实例化的贝叶斯网络

单个Plate 模型

刻画包含属性I(S)，G(S)的学生plate。偶对<Intelligence，Grade>的一个集合，每个偶对对应一个学生。

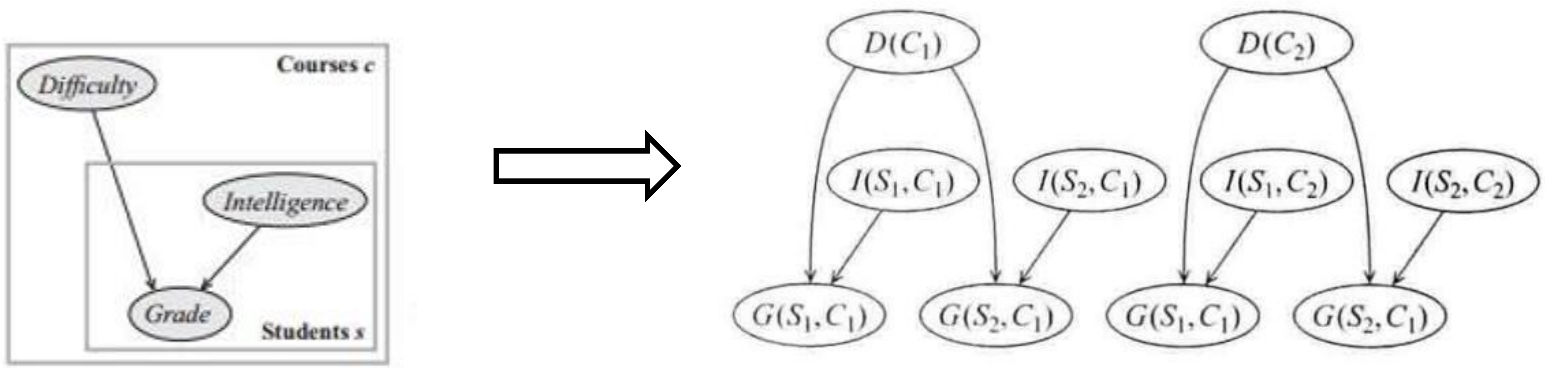


实例化这个模型得到的基础贝叶斯隐藏了一个假设所有的学生S有相同的CPD分布

Plate 模型和实例化的贝叶斯网络

有一门课程有多个学生，每个学生都有自己的分数，并且分数取决于课程的难易程度，引入第二种plate模型。

嵌套Plate 模型

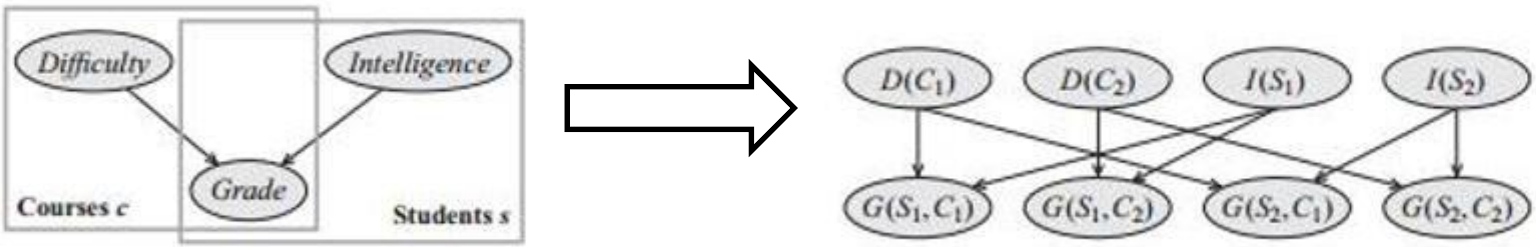


课程plate中嵌入了学生plate 变量difficulty包含在课程plate内 变量Intelligence 和 Grade 同时包含两个plate 中。

存在一个问题 intelligent不是关联学生 而是关联 学生和课程这个偶对中。如果一个学生选修了两门课程，就需要两个不同个变量来表示他的智力，而且这两个变量可以取不同的值。

Plate 模型和实例化的贝叶斯网络

相交Plate 模型



学生plate和课程plate相交 属性 intelligence 就只与学生plate关联 属性difficulty 只与课程plate关联，属性grade与偶对关联。

Plate 模型

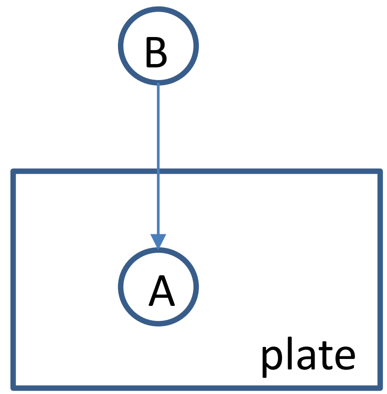
对每个具有参数名$U_1,\ldots U_k$的模板属性 $A \in \aleph$，plate模型M_{plate}定义了：

- 一个模板父节点的集合

$$Pa_A = \{B_1(U_1), \ldots B_L(U_L)\}$$

使得对每个$B_i(U_i)$,有 $U_i \subseteq \{U_1, \ldots U_k\}$ 。变量U_i是父节点B_i的参数名。

- 一个模板CPDP($A|Pa_A$)。



在plate的形式体系中，对象类型称为plate。在类中多个对象可以共享属性集和相同的概率模型的事实，是使用术语plate的基础，它意味着存在一大推相同的对象。

受限：
$A(U_1,\ldots U_k)$只能依赖于形如$B(U_{i1},\ldots U_{iL})$的属性，其中$U_{i1},\ldots U_{iL}$是$U_1,\ldots U_k$的一个自己。
这种约束极大的限制了我们在领域中描述对象之间概率依赖关系复杂网路的能力。

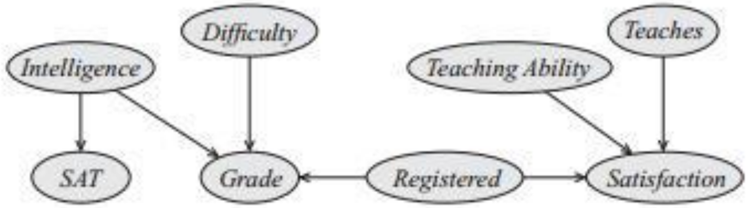
概率关系模型

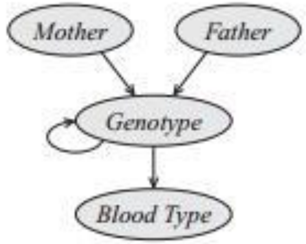
对于模板属性A,或有相依模型定义为由下列成分组成的一个元组：

- 父节点参数名 $\alpha(\mathrm{Pa}_A)$是满足 $\alpha(\mathrm{Pa}_A) \supseteq \alpha(\mathrm{A})$的类型化逻辑变量的一个元组
- 防护 Γ 是参数名 $\alpha(\mathrm{Pa}_A)$上根据模板属性集 Pa_A^Γ 定义的一个二值公式
- 模板父节点的一个集合 $Pa_A = \{\mathrm{B}_1(\mathrm{U}_1), \dots \mathrm{B}_L(\mathrm{U}_L)\}$

满足对每一个 $B_i(\mathrm{U}_i)$，有 $Ui \subseteq \alpha(\mathrm{Pa}_A)$

模板依赖模型MPRM的一个模板依赖图包含每个模板层属性A的一个节点，并且当类B的属性在 $Pa_A^\Gamma \cup Pa_A$ 中时，包含从B到A的一条有向边。





Thank you
Q and A