第九章 概率模型

第 26 讲 概率图模型

黄定江

DaSE @ ECNU
djhuang@dase.ecnu.edu.cn

1. 26.1 有向图与条件独立性
2. 26.2 无向图

26.1.1 引言

一个有向图是由一系列的节点及连接节点的有向边组成的。图1给出了一个有向图的例子。

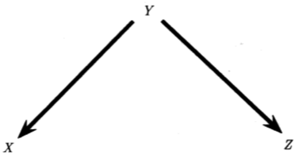

图 1: 节点集为 $V = X, Y, Z$ 且边集为 $E = (Y, X), (Y, Z)$ 的一个有向图

图在表示变量间的独立性关系方面是很有用处的，还可以用来代替反事实去表示因果关系。一个被赋予某种概率分布的有向图常被称为贝叶斯网络。频率学派或贝叶斯学派的方法都可以用来对有向图进行统计推断，所以贝叶斯网络这个说法是有歧义的。在进行关于有向非循环图 (DAGs) 的讨论之前，需要先讨论一下条件独立性。

26.1.1 条件独立性定义

定义 1

令 X, Y 和 Z 为随机变量。在给定 Z 的条件下，X 和 Y 称为条件独立的，记作 $X \perp\!\!\!\perp Y \mid Z$，如果下式对于所有的 x，y 和 z 均成立，

$$f_{X,Y|Z}(x, y \mid z) = f_{X|Z}(x \mid z) f_{Y|Z}(y \mid z)$$

一个等价的定义为

$$f(x \mid y, z) = f(x \mid z)$$

26.1.1 条件独立性的基本性质

定理 1

下列各蕴涵关系成立：

$$X \perp\!\!\!\perp Y \mid Z \Rightarrow Y \perp X \mid Z$$
$$X \perp Y \mid Z \text{ 且 } U = h(X) \Rightarrow U \perp Y \mid Z$$
$$X \perp Y \mid Z \text{ 且 } U = h(X) \Rightarrow X \perp\!\!\!\perp Y \mid (Z, U)$$
$$X \perp Y \mid Z \text{ 且 } X \perp\!\!\!\perp W \mid (Y, Z) \Rightarrow X \perp\!\!\!\perp (W, Y) \mid Z$$
$$X \perp Y \mid Z \text{ 且 } X \perp Z \mid Y \Rightarrow X \perp (Y, Z).$$

26.1.1 有向非循环图基本概念

- 一个有向图 $\mathcal{G}$ 是由节点集 V 及连接一对有序节点的边集 E 组成的。每个节点对应一个随机变量。若 $(X, Y) \in E$，则存在一条有向边从 X 指向 Y。见图1。
- 若一条有向边连接两个随机变量 X 和 Y(取任意一个方向)，就称 X 和 Y 是邻接的。若一条有向边从 X 指向 Y，则称 X 是 Y 的母节点，而 Y 是 X 的子节点。X 的所有母节点的集合记作 π_X 或 $\pi(X)$。两变量间的一条 c 是由一系列的同方向的有向边构成的，如下所示：

$$\mathbf{X} \longrightarrow \cdots \longrightarrow \mathbf{Y}$$

- 一个从 X 开始至 Y 结束的邻接节点的序列，但是忽略其有向边的方向性，就称该序列为一个无向路。图1中的序列 X, Y, Z 就是一个无向路。若存在一条有向路从 X 指向 Y (或 $X = Y$)，则称 X 是 Y 的祖节点。也可以说 Y 是 X 的后裔节点。

26.1.1 有向非循环图基本概念

- 如下形式的结构：

$$\boldsymbol{X} \longrightarrow \boldsymbol{Y} \longleftarrow \boldsymbol{Z}$$

称作在 Y 处相遇。不具有该种形式的结构称作不相遇，例如，

$$\boldsymbol{X} \longrightarrow \boldsymbol{Y} \longrightarrow \boldsymbol{Z}$$

相遇的性质是依赖于路的。在图1中，Y 是一个在路 X, Y, Z 上的相遇，但不是在路 X, Y, W 上的一个相遇。当指向相遇的变量不是邻接时，就说该相遇是无保护的。一条开始和结束都在同一个变量处的有向路是一个圈。若一个有向图没有圈，则它是非循环的。在这种情况下，称这种图为一个有向非循环图或 DAG。以后只考虑非循环图。

26.1.1 概率与 DAGs

令 $\mathcal{G}$ 为一个具有节点集 $V=(X_1,\cdots,X_k)$ 的 DAG。

定义 2

若 P 为 V 的分布，它的概率函数为 f，就说 P 是关于 $\mathcal{G}$ 是马尔可夫的，或称 $\mathcal{G}$ 表示 P，若下式成立：

$$f(v)=\prod_{i=1}^{k} f(x_i \mid \pi_i)$$

其中，T_i 为 X_i 的母节点。由 $\mathcal{G}$ 表示的分布集记为 $M(\mathcal{G})$。

26.1.1 马尔可夫举例

例 1

对于图2中的 *DAG* 来说，$\mathbb{P} \in M(\mathcal{G})$ 当且仅当其概率函数 f 具有以下形式：

$$f(x, y, z, w) = f(x)f(y)f(z \mid x, y)f(w \mid z)$$

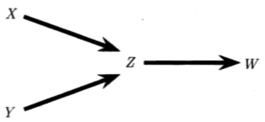

图 2: 另一个 DAG

26.1.1 马尔可夫条件成立定理

下述定理表明 $\mathbb{P} \in M(\mathcal{G})$ 当且仅当马尔可夫条件成立。粗略地讲，马尔可夫条件意味着每个变量 W 在给定其母节点的情况下与“过去”是独立的。

定理 2

一个分布 $\mathbb{P} \in M(\mathcal{G})$ 当且仅当下面的马尔可夫条件成立：对于每个变量 W，

$$W \perp \widetilde{W} \mid \pi_W$$

其中，$\widetilde{W}$ 表示除了 W 的母节点和后裔节点以外的所有其他变量。

26.1.1 独立性成立举例

例 2

考虑图3中的 *DAG*。在这种情况下，概率函数分解如下：

$$f(a,b,c,d,e)=f(a)f(b\mid a)f(c\mid a)f(d\mid b,c)f(e\mid d)$$

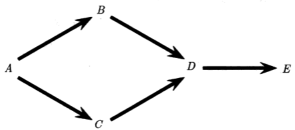

图 3: 另一个 DAG

马尔可夫条件意味着下面的独立性关系：

$$D\perp A\mid\{B,C\},\quad E\perp\{A,B,C\}\mid D \text{ 且} B\perp C\mid A.$$

26.1.1 DAGs 的估计

在 DAGs 中有两个首先要考虑的估计问题。
第一，给定一个 DAG$\mathcal{G}$ 和来自与 $\mathcal{G}$ 相符的分布为 f 的数据 $V_1, \cdots, V_n$，如何去估计 f?
第二，给定数据 $V_1, \cdots, V_n$，又如何去估计 $\mathcal{G}$?
第一个问题是一个纯粹的估计问题，而第二个问题则涉及到模型的选择。这些都是非常复杂的问题。这里仅简要介绍其主要思想。
通常，对于每个条件密度，人们常选择用某个参数模型 $f(x \mid \pi_x; \theta_x)$，则其似然函数为

$$\mathcal{L}(\theta) = \prod_{i=1}^{n} f(V_i; \theta) = \prod_{i=1}^{n} \prod_{j=1}^{m} f(X_{ij} \mid \pi_j; \theta_j)$$

其中，X_{ij} 是对于第 i 个数据点的 X_j 的值，θ_j 是第 j 个条件密度的参数。这样就可以通过极大似然方法来估计参数。

26.2.1 引言

无向图也可以像有向图一样来表示独立性关系。两者的主要差异是从图中读出独立性关系的规则不同。一个无向图 $G = (V, E)$ 由一个有限节点集 V 和由每对节点组成的边或 (弧) 集 E 所构成。节点对应着随机变量 $X, Y, Z, \cdots$ 而边被记作一些无序对。例如，$(X, Y) \in E$ 表示 X 和 Y 通过一条边连接起来。图4给出了一个无向图的例子。

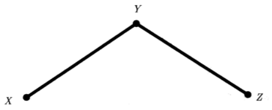

图 4: 节点集为 $V = \{X, Y, Z\}$ 的一个图。其边集为 $E = (Y, X), (Y, Z)$

26.2.1 无向图的基本概念

- 若两个节点之间存在一条边，则称这两个节点是邻接的，记作 $X \sim Y$. 在图4中，X 和 Y 是邻接的但是 X 和 Z 不是邻接的。若对每个 i 都有 $X_i - 1 \sim X$，则序列 $X_0, \cdots, X_n$ 称为一条路。在图4中，X, Y, Z 是一条路。若一个图中任意两个节点之间都存在一条边，则称这个图是完全的。一个子节点集 $U \in V$ 连同其边被称作一个子图。
- 设 A, B 和 C 是 V 的不同子集，若从 A 中的一个变量到 B 中的一个变量的路都相交于 C 中的一个变量, 就说 C 分离 A 和 B。在图5中，Y, W 和 Z 被 Z 分离。同时，W 和 Z 被 X, Y 分离。

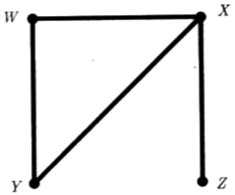

图 5: $\{Y, W\}$ 和 $\{Z\}$ 被 $\{X\}$ 分离。而且，W 和 Z 被 $\{X, Y\}$ 分离

26.2.1 概率与图

令 V 为具有分布 $\mathbb{P}$ 的随机变量集。构造一个图，其每个节点对应 V 中的每个变量。略去一对变量之间的边若它们在给定其余变量的条件下是独立的。即
X 和 Y 之间没有边 $\Leftrightarrow X \perp\!\!\!\perp Y \mid$ 其余变量，其中，“其余变量”表示除了 X 和 Y 之外的所有其他变量。这样的图称作成对马尔可夫图。如图6所示：

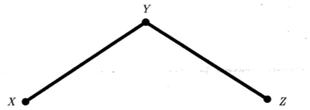

图 6: $X \perp\!\!\!\perp Y \mid Z$

图中暗含着一系列的成对条件独立性关系。这些关系可以推出其他的条件独立性关系。如何找到这些关系呢？幸运的是，也可以从图中直接读出这些其他的条件独立性关系，如下面的定理所述。

26.2.1 无向图定理

定理 3

令 $\mathcal{G} = (V, E)$ 是一个分布为 $\mathbb{P}$ 的成对马尔可夫图。令 A, B 和 C 为 V 的不相同的子集使得 C 分离 A 和 B，A II $B \mid C$。

定理3中的独立性条件被称作全局马尔可夫性质。将看到成对和全局马尔可夫性质是等价的。把这个问题表述得更确切些。给定一个图 $\mathcal{G}$，令 $M_{pair}(\mathcal{G})$ 表示满足成对马尔可夫性质的分布集，因此 $P \in M_{pair}(\mathcal{G})$，在分布 $\mathbb{P}$ 下，若 XII$Y \mid$ 其余变量当且仅当 X 和 Y 之间不存在边。令 $M_{global}(\mathcal{G})$ 为满足全局马尔可夫性质的分布集：则 $P \in M_{pair}(\mathcal{G})$，在分布 $\mathbb{P}$ 下，若 AB，A II $B \mid C$ 当且仅当 C 分离 A 和 B。

26.2.1 无向图

定理 4

令 $\mathcal{G}$ 为一个图，则 $M_{pair}(\mathcal{G}) = M_{global}(\mathcal{G})$。

例 3

由图7可知 $X \perp\!\!\!\perp Y$, $X \perp\!\!\!\perp Z$ 和 $X \perp\!\!\!\perp (Y, Z)$。

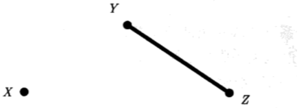

图 7: $X \perp\!\!\!\perp Y$

26.2.1 团和势

若一个图的变量集中的任意两个对应的节点都是邻接的，则称该集为一个团。若一个团任意增加一个节点后就不能成为团，则称之为一个极大团。一个势就是任意一个正函数。在特定的条件下，可以证明 $\mathbb{P}$ 关于 $\mathcal{G}$ 是马尔可夫的当且仅当其概率函数 f 可以写为

$$f(x)=\frac{\prod_{C\in\mathcal{C}}\psi_C\left(x_C\right)}{Z}$$

其中，$\mathcal{C}$ 是一个极大团集，ψ_C 是一个势，且

$$Z=\sum_x\prod_{C\in\mathcal{C}}\psi_C\left(x_C\right)$$

26.2.1 团和势举例

例 4

图4中的极大团是 $C = X, Y$ 和 $C = Y, Z$。因此，若 $\mathbb{P}$ 关于该图是马尔可夫的，则其概率函数可以写为

$$f(x, y, z) \propto \psi_1(x, y)\psi_2(y, z)$$

ψ_1 和 ψ_2 是某些正函数。

26.2.1 团和势举例

例 5

图8中的极大团为

$$\{X_1, X_2\},\quad \{X_1, X_3\},\quad \{X_2, X_4\},\quad \{X_3, X_5\},\quad \{X_2, X_5, X_6\}$$

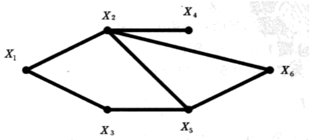

图 8: 该图的极大团为 $\{X_1, X_2\},\quad \{X_1, X_3\},\quad \{X_2, X_4\},\quad \{X_3, X_5\},\quad \{X_2, X_5, X_6\}$

因此可以把概率函数写为

$$f(x_1, x_2, x_3, x_4, x_5, x_6) \propto \psi_{12}(x_1, x_2)\,\psi_{13}(x_1, x_3)\,\psi_{24}(x_2, x_4) \times \psi_{35}(x_3, x_5)\,\psi_{256}(x_2, x_5, x_6)$$