# Bayesian 网中的独立关系*)

The Independence Relations in Bayesian Networks

王 飞 刘大有 卢奕男 薛万欣

(吉林大学计算机系 长春130023)

**Abstract** Bayesian networks are compact representation of joint probabilistic distribution. Independence is soul of Bayesian networks because it enables to save storage space, to reduce computational complexity and to simplify knowledge acquisition and modeling. In this paper, we discuss three kinds of independences in Bayesian networks: conditional independence, context-specific independence and causal influence independence.

**Keywords** Bayesian networks, Conditional independence, Context-specific independence, Casual influence independence

## 1 引言

在不确定性下进行推理和做出决策的能力是智能行为的基础。在过去几十年中，大量研究人员尝试了多种方法研究不确定性知识的表示和运用，其中有证据理论模型、确定性因子、PROSPECTER 模型、模糊集理论以及近年来逐渐成为主流的 Bayesian 网等。Bayesian 网是图形表示方式和概率知识的有机结合，它揭示领域对象的内在结构，是复杂联合概率分布的紧凑表示方式。其坚实的理论基础、知识结构的自然表述方式、灵活的推理能力、方便的决策机制及有效的学习能力使其应用范围越来越广泛。

首先说明文中不同变量的含义。如果没有特殊说明，文中大写字母 X、Y、Z 表示随机变量；小写字母 $x$，$y$，$z$ 表示随机变量的取值；$val(X)$表示随机变量 $X$ 的取值范围；黑体 $\mathbf{X}$，$\mathbf{Y}$，$\mathbf{Z}$ 表示随机变量的集合。

Bayesian 网是一个有向非循环图，图中节点与知识领域的随机变量一一对应（文中不区分节点与变量）；网中的有向弧表示变量间的因果关系，从节点 X 到节点 Y 有向弧的直观含义是 X 对 Y 有直接的因果影响；因果影响不是绝对的，影响的强度或者说不确定性由条件概率表示，每个节点有一个条件概率表，定量描述其所有父亲节点对于该节点的作用效果。

给定 Bayesian 网 $T$，则存在一个变量集合 $\mathbf{X}=\{X_1,X_2,\cdots,X_n\}$上的联合概率分布 $P$，$\mathbf{X}$ 的变量和 $T$ 中的节点一一对应，并且 $P$ 中存在递归展开式：

$$p(X_1,\cdots,X_n)=\prod_i p(X_i|\pi_{X_i}) \quad (1)$$

其中 $\pi_{X_i}$是 $X_i$ 在 $T$ 中的父亲节点集合。这个递归展开式意味着对变量 $X_i$，给定其父亲集合 $\pi_{X_i}$，$X_i$ 独立于除 $\pi_{X_i}$外的所有先驱节点。联合概率分布可以解答关于领域的任何查询。例如，已知关于 $n$ 个变量的联合概率分布 $p(X_1,X_2,\cdots,X_n)$，则 $p(x_2,\cdots,x_n)=\sum_{x_1}p(x_1,x_2,\cdots,x_n)$，而 $p(X|Y)=\frac{p(X,Y)}{p(Y)}$，利用这两个式子可以计算任何概率。Bayesian 网是联合概率分布的简化表示形式，可以计算变量空间的任意概率值。当变量数目很大时，运用联合概率分布进行计算通常是不可行的，概率数是变量数目的指数幂，计算量难以承受。Bayesian 网利用独立关系解决了这个难题。Bayesian 网中存在三种独立关系：条件独立、上下文独立及因果影响独立。三种独立关系旨在把联合概率分布分解成颗粒更小的因式，从而达到节省存储空间，简化知识获取和领域建模过程，降低推理过程中计算复杂性的目的，因此可以说独立关系是 Bayesian 网的灵魂。

## 2 Bayesian 网中的独立关系

### 2.1 条件独立关系

Bayesian 网的图形结构本身已经表达出条件独立关系，每个节点在已知其父亲节点的条件下独立于所

*)本项研究得到国家863高技术项目、高等学校博士学科点专项科研基金和教育部符号计算与知识工程重点实验室的资助。王飞 博士生。刘大有 教授，博士导师。卢奕男 博士生，薛万欣 博士生。研究方向：不确定性推理，分布式人工智能，神经网络与遗传算法等。

有其余的非子孙节点。利用条件独立，Bayesian 网把联合概率分解成若干个条件概率的乘积，每个条件概率涉及的数量数目较少，从而达到粒化的目的。d-分离是一种可以判断任意两个变量间是否存在条件独立的图形判别方法。

**定义1（阻塞）** 一条路径被节点集 E 阻塞，是指在路径上存在一个节点 z 满足下面三种情形之一：1）z∈E，并且路径中有一条有向弧指向 z，另一条有向弧源自 z；2）z∈E，并且路径中有两条有向弧源自 z；3）z 及 z 的所有后继节点都不在 E 中，并且路径中有两条有向弧指向 z。

**定义2（d-分离）** X、Y、E 是有向非循环图 T 中三个互不相交的节点集，说 E d-分离 X 和 Y，记作 $I(X,Y|E)_T$，当且仅当 X 中任一节点和 Y 中任一节点间的任一条路径都被节点集 E 阻塞。

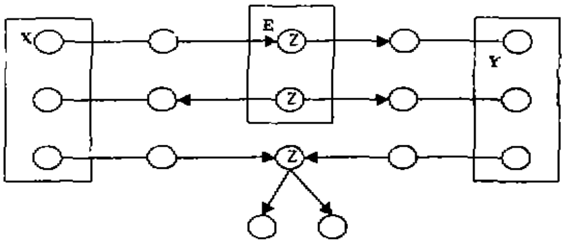


图1 图形方式表示 d-分离

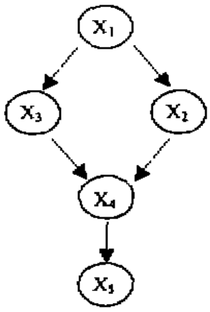


图2 表示五个变量间因果影响的 Bayesian 网

例如，在图2中，$X=\{X_2\}$ 和 $Y=\{X_3\}$ 被 $E=\{X_1\}$ $d$-分离，路径 $X_2 \leftarrow X_1 \rightarrow X_3$ 被 $X_1 \in E$ 阻塞，同时另一条路径 $X_2 \leftarrow X_4 \rightarrow X_3$ 也被阻塞，因为 $X_4$ 和它的所有后代节点都不在集合 $E$ 中。因此 $I(X_2,X_3|X_1)$ 存在于 $T$ 中。图2中，$X$ 和 $Y$ 未被 $Z'=\{X_1,X_5\}$ $d$-分离，一旦 $X_5$ 已知，$X$，$Y$ 就不再是独立的。

对于两个随机变量集合 X、Y，给定节点集合 E 时，X 和 Y 是条件独立的，但可能存在随机变量集 F，X 和 Y 在 E∪F 条件下不是独立的。尤其两个变量 X、Y 可能是相互独立的 $p(X|Y)=p(X)$，给定变量 $E$、$X$、$Y$ 在 $E$ 条件下不独立，即 $p(X|Y,E)\neq P(X|E)$。

可以证明，d-分离标准是有向非循环图中判断条件独立关系的充要条件，即递归展开式（1）蕴涵的条件独立和满足 d-分离标准的三元组 I(X,Y|E) 之间是一一对应的。并且，d-分离判别方法可以在线性于有向弧数目的时间内完成。

**定理1** 给定 Bayesian 网 T，其确定的唯一联合概率分布 P，X、Y、Z 是 T 中三个节点集，$I(X,Y|Z)$ 当且仅当 P(X|Y,Z)=P(X|Z)。

### 2.2 上下文独立关系

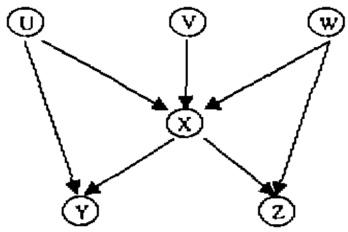


图3 上下文独立

表1

| U | V | W | P(x) |
|---|---|---|---|
| t | t | t | $P_1$ |
| t | t | f | $P_1$ |
| t | f | t | $P_1$ |
| t | f | f | $P_1$ |
| f | t | t | $P_2$ |
| f | t | f | $P_2$ |
| f | f | t | $P_3$ |
| f | f | f | $P_4$ |

Bayesian 网的每个节点 $X$ 都带有一张条件概率表，在 $\pi_X$ 的每种取值情况下给出 $X$ 每一种取值的概率，即 $p(X|\pi_X)$。条件概率表一方面需要的条件概率数目是父亲节点数目的指数幂，另一方面无法捕捉条件概率分布的某些规律。例如，在表1中，在 U=t 时，不需要考虑 V，W 的取值，这时最多只需要5个条件概率值，而不是8个。在一些知名的 Bayesian 网产品中已经注意到这类问题，如微软的 Bayesian 网建造工具，其知识获取接口引入特殊机制，使用户容易给出各节点的条件概率，减少工作量。

表1中，U=f 时，X 与 V、W 相关；U=t 时，X 与 V、W 独立，这是另一种独立关系——上下文独立，它是某些变量取特定值之后，其余变量之间存在的独立关系。Bayesian 网的结构可以表达出条件独立，对上下文独立无能为力。但是通过有效地表示条件概率，如摒弃概率表的形式，采用条件概率树，可以利用上下文独立关系。实质上，条件独立关系可以看作是一种极特殊的上下文独立关系，I(X,Y|Z) 表示一旦已知 Z，无论

其值是什么，X，Y 条件独立，

**定义3**(上下文独立)　X，Z，Y，C 是两两互不相交的随机变量集合，给定 Z 及上下文 $c\in val(C)$，X，Y 是上下文独立的，记作 $I_c(X,Y|Z,c)$，如果 $p(X|Z,c,Y)=p(X|Z,c)$，其中 $p(Y,Z,c)>0$。

由定义可看出，当 C 为空时，上下文独立变成特殊的条件独立。

d-分离提供了一种判断全局任意两个变量集是否存在条件独立关系的方法；同样，存在判断全局两个变量集是否存在上下文独立关系的方法。

**定义4**　给定 Bayesian 网 T 和上下文 $c$，如果 $I_c(X,Y|\pi_X\cap c)$，则称从 Y 到 X 的有向弧是无意义的。在上下文 $c$ 条件下，删除 Bayesian 网 T 中的无意义弧得到的结果网络定义为 T(c)。在 T(c)中，如果 X 和 Y 被 Z∪C d-分离，则称给定上下文 $c$ 及变量表 Z，X 和 Y 被上下文分离。

$I_c(X,Y|\pi_X\cap c)$是局部的上下文独立关系，可以从 X 节点的条件概率表中得出；使用局部的 $I_c$ 论断删除无意义弧把 Bayesian 网 T 简化成 T(c)；在 T(c)利用图形判别方法 d-分离就可以判断出任意两个变量集是否被上下文分离，d-分离是判断条件独立的充要条件，上下文分离是判断上下文独立的充要条件。

### 2.3　因果影响独立关系

Bayesian 网中的有向弧是一种因果关系，表示父亲节点对儿子节点的直接影响，但是没有对儿子如何依赖父亲集作出约束，在最坏情况下，需要给出的条件概率数目是父亲节点数目的指数幂。一些情况下，父亲节点(原因)之间相互合作，对儿子节点(结果)有一个共同的影响。但是，很多情况下，各个父亲独自对儿子起作用，父亲节点之间没有合作，我们说父亲节点对儿子节点的影响是因果独立的。

**定义5**(因果影响独立)　我们说节点 X 的各个父亲 $\pi_{X_1},\cdots,\pi_{X_m}$ 对于 X 是因果影响独立的，如果对应于 $\pi_{X_1},\cdots,\pi_{X_m}$ 存在和 X 有共同取值范围的随机变量 $\varepsilon_1,\cdots,\varepsilon_m$，并且下面两个条件成立：

(1)对于每个 i，$\varepsilon_i$ 仅仅在概率意义上依赖于 $\pi_{X_i}$、在 $\pi_{X_i}$ 条件下独立于所有其它的 $\pi_{X_j}$ 及 $\varepsilon_j(j\neq i)$，即 $I(\varepsilon_i,\{\pi_{X_1},\cdots,\pi_{X_{i-1}},\pi_{X_{i+1}},\cdots,\pi_{X_m},\varepsilon_1,\cdots,\varepsilon_{i-1},\varepsilon_{i+1},\cdots,\varepsilon_m\}|\pi_{X_i})$。

(2)存在一个定义域是 X 的取值范围，且具有交换律和结合律的二元运算符 *，使得 $X=\varepsilon_1*\varepsilon_2*\cdots*\varepsilon_m$ 成立。* 称作是 X 的基本结合运算符。

我们把 $\varepsilon_i$ 称作是 $\pi_{X_i}$ 对 X 的贡献。粗略地讲，有共同作用结果的多个原因是因果影响独立的，如果每个原因的各自贡献是独立的，所有原因对结果的影响是各自贡献的简单组合，因果影响独立大大降低每个节点所需的条件概率数目，从指数级降到线性级，当父亲节点很多时，降幅是十分巨大的，这里的因果影响独立涵盖了一般的因果影响独立模式，如噪音-或模式、噪音-与模式、噪音-最大模式、噪音-加模式等。下面给出三个不同模式因果影响独立的例子。

**例1**(噪音-加模式)　买彩票可能赚钱，也可能赔钱，买不同种类的彩票赚钱或赔钱是相互独立的，如买体育彩票没中彩赔钱不影响买福利彩票中彩赚钱。让 $\pi_{X_1},\cdots,\pi_{X_m}$ 表示买不同种类彩票所花的钱，$\varepsilon_1,\cdots,\varepsilon_m$ 表示彩票各自赚赔钱的数目，$\varepsilon_i$ 仅仅依赖于 $\pi_{X_i}$，独立于其余所有的 $\pi_{X_j}$ 及 $\varepsilon_j(j\neq i)$。$X$ 表示买彩票引起财富变化的总数，则有 $X=\varepsilon_1+\varepsilon_2+\cdots+\varepsilon_m$。买不同种类的彩票对于钱总数的改变是因果独立的，X 的基本组合运算符是数值加运算，

**例2**(噪音-或模式)　一个盗窃报警器装有 m 个不同类型的感知器，一个感知器被触发，报警器就会响铃报警。用 $\varepsilon_i$ 表示第 i 个感知器的状态，感知器类型不同，是独立差错的，报警 $alarm=\varepsilon_1\vee\varepsilon_2\vee\cdots\vee\varepsilon_m$。因此每个感知器是否引起报警器报警是因果影响独立的。基本组合运算符是逻辑或运算。

**例3**　某所大学对每个教师从教学、科研、其它三方面进行综合评定，目的是决定是否续签工作合同。工作合同可能不续签、或续签不加薪、或续签加少量薪水、或加一倍薪水，依据是上面三个方向的评定结果：在一个方面不合格、所有方向都合格、在一个方面表现优秀、在至少两个方面表现优秀。用 $\pi_{X_1},\pi_{X_2},\pi_{X_3}$ 表示一个教师在教学、科研、其它三方面得到的分数，$\varepsilon_i$ 表示在第 $i$ 个方面得到的评定等级，取值0—不合格、1—合格、2—优秀。$\varepsilon_i$ 依赖于 $\pi_{X_i}$，并且独立于其余的 $\pi_{X_j}$ 及 $\varepsilon_j(j\neq i)$。$X$ 表示合同的结果，取值范围0—不续签、1—续签不加薪、2—加少量薪水、3—加一倍薪水，则 $X=\varepsilon_1*\varepsilon_2*\varepsilon_3$。基本运算符 * 的运算规则如表2所示。可以得出结论，教师在三个方面的表现对于合同的结果是因果影响独立的。

表2

| | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 1 | 2 | 3 |
| 2 | 0 | 2 | 3 | 3 |
| 3 | 0 | 3 | 3 | 3 |

## 3　独立关系的作用

独立关系在知识表示、推理、学习方面起到的简化

作用使得 Bayesian 网作为一种内涵方法，计算复杂性大大降低，可用性和实用性大大增强。

独立关系在节省存储空间方面的作用是显而易见的。如图4，八个变量的联合概率分布以 $p(X_1,\cdots,X_8)$ 的形式存储，不加以任何简化，需要$2^8-1=255$个概率值。利用独立关系粒化联合概率分布，首先运用条件独立和因果影响独立得到下式

$$p(X_1,\cdots,X_8)=p(X_8|X_3,X_7)p(X_7|X_4,X_5,X_6)p(X_6|X_1)p(X_5|X_2)p(X_4|X_2)p(X_3|X_1)p(X_2|X_1)p(X_1)$$
$$=p(X_8|X_3,X_7)p(X_7|X_4)p(X_7|X_5)p(X_7|X_6)p(X_6|X_1)p(X_5|X_2)p(X_4|X_2)p(X_3|X_1)p(X_2|X_1)p(X_1)$$

再利用上下文独立关系，其中 $p(X_8|X_3,X_7)=(p(X_8|x_3,X_7),p(X_8|\neg x_3))$

简化后的联合概率分布表示只需$3+2+2+2+2+2+2+2+2+1=20$个概率值，存储空间节省20倍以上。随着变量数目的增多，存储空间的节约更加可观。

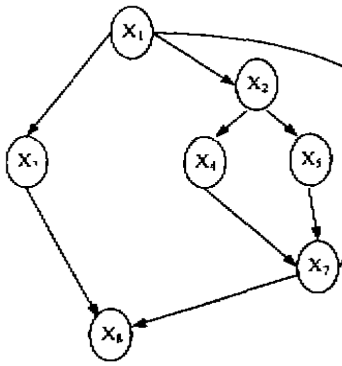


图 4

图4中变量都是二值变量，$X_4$、$X_5$、$X_6$对于 $X_7$的因果影响是相互独立的，即

$p(X_7|X_4,X_5,X_6)=p(X_7|X_4)p(X_7|X_5)p(X_7|X_6)X_8$ 在 $X_3=false$ 时和 $X_7$上下文独立，即 $p(X_8|X_3=false,X_7)=p(X_8|X_3=false)$。

Bayesian 网推理是概率推理，是信念更新的过程。当新证据出现时，一些变量的后验概率发生变化，有时人们说节点的信念发生变化，因为人们一般认为这里的信念指后验概率。简而言之，Bayesian 网推理是在给定模型中计算目标变量的后验概率，这个概率不能从网络中直接读出，必须通过计算。Cooper(1990)已经证明对应于一般问题的 Bayesian 网，后验概率的计算是 NP 难度的。幸运的是，当所有变量都是离散变量时，领域结构可以大大简化计算。一些学者利用独立关系，研究出一些关于离散变量的 Bayesian 网概率推理算法。在每种具体的算法中，独立关系的表示形式不同，应用独立关系降低推理复杂度的方法不同。

本文以逐个变量求和的推理算法为例，说明独立关系如何降低推理复杂度。因为联合概率分布的表示形式是若干个因式，即先验概率和条件概率的集合，所以需要定义如何从概率集合 S 中对一个变量 V 的概率分布求和。具体方法是：(1)从 S 中删除所有包含 V 的局部概率分布，(2)在包含 V 的所有概率分布中对 V 的概率分布求和，(3)求和结果加入到概率集 S 中。

例如，在图4中对变量 $X_1$的概率分布求和，首先把 $p(X_6|X_1)$，$p(X_3|X_1)$，$p(X_2|X_1)$，$p(X_1)$从图的联合概率分布表示中删去，然后计算

$$\varphi_{X_1}(X_2,X_3,X_6)=\sum_{x_1}p(X_6|x_1)p(X_3|x_1)p(X_2|x_1)p(x_1) \quad (2)$$

最后把 $\varphi_{X_1}(X_2,X_3,X_6)$加入到联合概率分布表示中。图4的联合概率分布表示变为

$\{\varphi_{X_1}(X_2,X_3,X_6),P(X_8|x_3,X_7),P(X_8|\neg x_3),p(X_7|X_4),p(X_7|X_5),p(X_7|X_6),p(X_5|X_2),p(X_4|X_2)\}$

一般情况下，从联合概率分布的因式表示形式中对一个变量的概率分布求和比从没有任何简化的联合概率分布表示涉及的数学计算要少得多。式(2)说明了从图4中对变量 $X_1$的概率分布求和需要的所有数学计算，仅仅和变量 $X_1$，$X_2$，$X_3$，$X_6$相关，而不需要执行运算$\sum_{x_1}p(x_1,X_2,X_3,X_4,X_5,X_6,X_7,X_8)$，它和图4中所有的变量相关。这就是运用独立关系粒化联合概率分布表示能够降低计算复杂度的原因。

## 参考文献


1 Cooper G. Computational complexity of probabilistic inference using Bayesian belief networks(Research note). Artificial Intelligence，1990，42：393～405

2 Friedman J. Introduction to computational learning and statistical prediction：[Technical report]. Department of Statistics，Stanford University. 1995

3 Boutilier C，et al. Context-specific independence in Bayesian networks. UAI'96，Morgan Kaufmann. 115～123

4 Pearl J. Fusion，propagation，and structuring in belief networks. Artificial Intelligence，1986，29：241～288

5 Pearl J. Probabilistic Reasoning in Intelligent Systems. Networks of Plausible Inference. Morgan Kaufmann，San Mateo，CA 1988

6 Pearl J. Belief networks revisited. Artificial Intelligence，1993，59：49～56

7 Pearl J. Causal diagrams for empirical research. Biometrika，1995，82：669～710

8 Zhang N. Inference in Bayesian networks：The role of Context-Specific Independence：[Tech Report HKUST-CS98-09]. 1998

9 Zhang N，Poole D. A simple approach to Bayesian network computations. In：Proc. of the Tenth Canadian Conf. on Artificial Intelligence，1994. 171～178

10 Zhang N，Poole D. Exploiting causal independence in Bayesian network inference. Journal of Artificial Intelligence Research，1996，5：301～328

11 Zhang N，Poole D. On the role of Context-Specific Independence in Probabilistic inference. In UAI'98

12 Zhang T，Ramakrishaan R，Livny M. Birch：An efficient data clustering method for very large databases. In：Proc. of the Fifteenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Syetems. ACM，1996