

文章编号:1001-9081(2005)01-0001-03

基于信息论的 Bayesian 网络结构学习算法研究

聂文广, 刘惟一, 杨运涛, 杨 明
(云南大学 计算机科学系, 云南 昆明 650091)
(nqlnwg@sohu.com)

摘 要: Bayesian 网是一种进行不确定性推理的有力工具, 它结合图型理论和概率理论, 可以方便地表示和计算我们感兴趣的事件概率, 同时也是对实体之间依赖关系提供了一种紧凑、直观、有效的图形表示。文中基于信息论中测试信息独立理论, 对 Bayesian 网中各结点进行条件独立(CI)测试, 以发现各结点的条件依赖关系, 并通过计算结点之间的互相依赖度以发现 Bayesian 网边的方向, 从而构造 Bayesian 网结构, 算法的计算复杂度只需要进行 $O(N^2)$ 次 CI 测试。

关键词: Bayesian 网络; 结构学习; 条件独立性; 条件互信息; 条件依赖度

中图分类号: TP393.02 **文献标识码:** A

Algorithm of Bayesian network structural learning based on information theory

NIE Wen-guang, LIU Wei-yi, YANG Yun-tao, YANG Ming

(Department of Computer Science, Yunnan University, Kunming Yunnan 650091, China)

Abstract: Bayesian network is a forceful tool to practise inference of uncertainty. It combines graphic theories and probability ones, which can conveniently express and calculate the probability of interesting events and at the same time provide a compact, visual and effective graphic expression for the dependant relationship among the entities. On the basis of testing information independence theory, the test of CI(conditional independence) was carried out on all the joints in the Bayesian network to find out the conditionally dependant relations among them. Then an effective algorithm of Bayesian network structural learning was worked out, which only needed CI testing of $O(N^2)$ times.

Key words: Bayesian network; structural learning; Conditional Independence(CI); conditionally mutual information; degree of conditional independence

0 引言

Bayesian 网是一种进行不确定性推理的有力工具, 它结合图型理论和概率理论, 可以方便地表示和计算我们感兴趣的事件概率, 同时也是对实体之间依赖关系提供了一种紧凑、直观、有效的图形表示。近十年来, 随着 Bayesian 网的研究逐渐成熟, Bayesian 网已被广泛应用于人工智能、专家系统以及数据挖掘等领域。

Bayesian 网络的结构构造方法一般可以分为三种方法: 1) 从专家知识获得; 2) 对已有的模型进行提炼; 3) 从大量数据中学习 Bayesian 网。由于前两种方法带有较大的主观性和局限性, 所以第 3 种方法已成为人们进行 Bayesian 网络构造的主要研究对象。

从大量数据中学习 Bayesian 网的构造方法又可以分为两类:

1) 基于打分搜索的方法(也叫贝叶斯方法)。这种方法就是对所有可能的 Bayesian 网络结构进行搜索, 使用某种打分函数对所有可能的结构进行打分, 从而寻找一个与数据拟合得最好的网络结构。常用打分函数方法有基于求最大后验结构概率算法、基于最小描述长度方法、最大熵方法等等。这

种方法从理论上来说是可以找到一个较好的 Bayesian 网络的结构, 但是如何从 n 个结点中找出适合条件独立的顺序, 需要对 $n!$ 种可能的网络结构进行打分计算, 这是一个组合爆炸的问题, 实质上已证明这是一个 NP-难题, 为了降低搜索空间, 很多研究在使用这种方法时都对其做了某些限定, 如限定结点的个数、假定结点有序、根据专家知识采用某些启发式的搜索规则等等。这样就大大降低了计算的难度, 使得打分搜索方法变得可行, 但另一方面由于这些限定的使用, 又使得这种打分搜索的学习方法不一定能找到最好的网络结构。

2) 基于依赖分析的方法(也叫非贝叶斯方法)。一个 Bayesian 网的结构本质上就是表示一个系统中的属性之间的某些依赖关系(有向无环图), 所以它又称为因果关系网或信度网, 基于这种理论基础, 非贝叶斯方法就是首先发现系统中属性之间的某些依赖关系, 而后根据这些依赖关系来建立 Bayesian 网。常用的基于依赖分析的方法有: 用 Polytrees 来表示概率网的方法、从完全图删除边的方法等等。这种方法需要对条件集进行指数次的 CI 测试, 以发现依赖关系。但是这种方法需要有完整的、大的样本数据集, 当样本数据集越大时, 测试的结果越可靠, 否则这种测试的结果不一定可靠; 其次它需要进行指数级的 CI 测试, 当结点集较大时, 它的计算效率低,

收稿日期: 2004-06-14; 修订日期: 2004-12-09 基金项目: 国家自然科学基金资助项目(60263003); 云南省自然科学基金资助项目(2002F0011M); 科学院智能信息处理开发实验室课题(IIP2002-2)

作者简介: 聂文广(1962-), 男, 副教授, 硕士研究生, 主要研究方向: 数据与知识工程; 刘惟一(1950-), 男, 教授, 博士生导师, 主要研究方向: 数据与知识工程。

所以大多数此类算法都假设结点有序。对于稀疏网络和具有较大样本数据集的系统来说,这种方法是非常有效的。

本文基于信息论中测试信息独立理论,对 Bayesian 网中各结点进行条件独立(CI)测试,以发现各结点的条件依赖关系,并通过计算结点之间的相互依赖度以发现 Bayesian 网的方向,而不需要假设结点的有序。

1 Bayesian 网与依赖模型^[4,5]

定义 1: Bayesian 网是一个有向无环图(DAG),它可以表示为一个三元组 $G = (N, E, P)$ 。其中 N 是一组结点的集合, $N = \{x_1, x_2, \dots, x_n\}$, 每个结点代表一个变量(属性)。 E 是一组有向边的集合, $E = \{ \langle x_i, x_j \rangle \mid x_i \neq x_j \text{ 并且 } x_i, x_j \in N \}$, 每条边 $\langle x_i, x_j \rangle$ 表示 x_i, x_j 具有依赖关系 $x_i \rightarrow x_j$ 。 P 是一组条件概率的集合, $P = \{p(x_i \mid \pi_i) \mid p(x_i \mid \pi_i) \text{ 表示 } x_i \text{ 的父结点集 } \pi_i \text{ 对 } x_i \text{ 的影响}\}$ 。

设某个系统有 N 个属性结点 $\{x_1, x_2, \dots, x_n\}$ 由联合概率公式可得:

$$p(x_1, x_2, \dots, x_n) = \prod_{i=1}^n p(x_i \mid x_1, x_2, \dots, x_{i-1}) \quad (1)$$

现在假如对每个结点 x_i , 已知存在某个子集 $\pi_i \subseteq \{x_1, x_2, \dots, x_{i-1}\}$, 在给出 π_i 时, x_i 和 $\{x_1, x_2, \dots, x_{i-1}\} - \pi_i$ 是条件独立的, 即有:

$$p(x_i \mid x_1, x_2, \dots, x_{i-1}) = p(x_i \mid \pi_i) \quad (2)$$

由(2)式,(1)可等价:

$$p(x_1, x_2, \dots, x_n) = \prod_{i=1}^n p(x_i \mid \pi_i) \quad (3)$$

由(3)式可知, Bayesian 网实质上就是一个联合概率分布 $p(x_1, x_2, \dots, x_n)$ 所有条件独立性的图形化表示, 其中 π_i 为 x_i 父亲结点集。

定义 2: 依赖模型 M 定义为一组条件独立的集合, 设 X, Y, Z 是全集 U 的三个不相交的子集, $M = \{I(X, Z, Y)\}$ 。其中 $I(X, Z, Y)$ 表示在给定 Z 的条件下, X 独立于 Y , 即: $p(X \mid Y, Z) = p(X \mid Z)$ 和 $p(Y \mid X, Z) = p(Y \mid Z)$ 。

定理 1^[1]: 在依赖模型 M 中, 设 X, Y, Z 是全集 U 的三个不相交的子集, 条件独立 $I(X, Z, Y)$ 满足对称性、分解律、减缩律、交换律、弱归并律等性质。

从依赖模型 M 出发来构造 Bayesian 网的结构是我们最终的目标, 那么给出一个 M , 如何用一个 DAG 来表示 M 呢? 我们可以利用 DAG 中 d-分离的概念来表示依赖模型中的条件独立。

定义 3: (V-结构) 设 a, b, c 是有向无环图 G 中三个不同的结点, 如果满足 $a \rightarrow b \in G$ 和 $c \rightarrow b \in G$, 且 a 和 c 之间在 G 中不存在有向边, 我们就称三元组 (a, b, c) 为图 G 的一个 V-结构, b 称为汇聚结点。

定义 4: (d-分离) 在有向无环图 G 中, X, Y, Z 是 U 上三个不相交的子集, 如果从 X 中一个结点到 Y 中的一个结点的所有路径之间, 存在结点 E 满足以下条件之一: 1) E 不是一个汇聚结点, 且 E 属于 Z 中; 2) E 是一个汇聚结点, 且 E 或任何 E 的子结点都不属于 Z 中。则称 X 和 Y 对于 Z 为 d-分离, 记为 $\langle X \mid Z \mid Y \rangle_G$ 。

这样就可以用 $\langle X \mid Z \mid Y \rangle_G$ 来表示依赖模型中条件独立信息 $I(X, Z, Y)$, 从而得到了一个依赖模型的图形化表示方

式, 但是这种图形化的表示能否正确完全地反映依赖模型 M 中的所有条件独立信息呢? 是不是所要求的最优 Bayesian 网呢? 下面我们继续用 I-图、P-图、D-图的概念来进行衡量。

定义 5: 设 M 为概率依赖模型, $\langle X \mid Z \mid Y \rangle_M$, 表示依赖模型 M 所蕴涵的依赖关系 $I(X, Z, Y)$, 定义:

1) 当 $\langle X \mid Z \mid Y \rangle_M = \langle X \mid Z \mid Y \rangle_G$, 则称 G 是 M 的依赖图, 记 D-图;

2) 当 $\langle X \mid Z \mid Y \rangle_G = \langle X \mid Z \mid Y \rangle_M$, 则称 G 是 M 的独立图, 记 I-图;

3) 当 $\langle X \mid Z \mid Y \rangle_M \subseteq \langle X \mid Z \mid Y \rangle_G$, 则称 G 是 M 的理想图, 记 P-图。

由定义可知, 对于任一依赖模型 M , 我们总想找到它所对应的 P-图, 但实际上这并不是总能办到的, 所以我们往往从 I-图着手, 找到一个对应 M 的最小 I-图。

定义 6: 设一个有向无环图 G 是 M 的一个 I-图, 若删除 G 中的任何一条边后, 使得 G 不再是 M 的 I-图, 则称 G 为 M 的最小 I-图。

显然, 最小 I-图能够最多地表示依赖模型 M 中的依赖关系, 也即是我们所希望发现的 Bayesian 网。

定理 2^[1]: 满足对称性、分布性、交换律和弱归并律的依赖模型 M , 从完全图中删除所有条件独立性成立的边, 则产生一个唯一的最小 I-图。

2 信息论概念^[6]

信息熵是对信息不确定性的一种度量。从信息论角度来看, 信息就是用来消除不确定性的东西, 信息的载体称为消息, 含有信息的消息集合称为信源, 信源的信息熵, 就是信源提供的整个信息的总体度量。所以如果消息消除的不确定性越大, 信源的信息熵就越小, 信息间的相互依赖性就越大; 反之信息间的相互独立性就越大。具体概念做如下定义:

定义 7: 设 X, Y, Z 为三个不相交的变量集, 称:

$I(X, Y) = \sum_{i=1}^r \sum_{j=1}^q p(x_i, y_j) \log \left(\frac{p(x_i, y_j)}{p(x_i)p(y_j)} \right)$ 为 X, Y 的互信息。

$$I(X, Y \mid Z) =$$

$$\sum_{i=1}^r \sum_{j=1}^q \sum_{k=1}^s p(x_i, y_j, z_k) \log \left(\frac{p(x_i, y_j \mid z_k)}{p(x_i \mid z_k)p(y_j \mid z_k)} \right)$$

为给定 Z 的条件下, X 和 Y 的互信息(条件互信息)。

其中 r, q, s 为 X, Y, Z 的状态个数, $P(x_i, y_j, z_k)$ 为 (X, Y, Z) 的状态为 (x_i, y_j, z_k) 时的概率。

定理 3^[7]: 互信息 $I(X, Y)$ 和 $I(X, Y \mid Z)$ 具有如下性质:

1) 对称性, 即 $I(X, Y) = I(Y, X)$ 和 $I(X, Y \mid Z) = I(Y, X \mid Z)$;

2) 非负性, 即 $I(X, Y) \geq 0$ 和 $I(X, Y \mid Z) \geq 0$ 。而且, 当且仅当 X 和 Y 条件独立时有 $I(X, Y) = 0$ 。同理, 有当且仅当在给定条件 Z , X 和 Y 条件独立时有 $I(X, Y \mid Z) = 0$ 。

3 基于信息论的 Bayesian 网的算法

首先给出这个算法的所需要的一些假设:

假设 1: 我们给定的样本数据集 D 是完整的;

假设 2: 所有的变量取值均为离散性, 如取值连续的话可先进行离散化;

假设3:假设每个结点 X 对其所有父结点 Pa_1, Pa_2, \dots, Pa_m 的依赖是互相独立的。

第一步:构造完全有向图

定义8:设一个系统含有 N 个变量 $\{X_1, X_2, \dots, X_n\}$,完全有向图 $P_c = \{ \langle X_i, X_j \rangle \mid \text{其中 } i, j = 1, 2, \dots, n \text{ 且 } i \neq j, \langle X_i, X_j \rangle \text{ 表示 } X_i \text{ 与 } X_j \text{ 有因果关系 } X_i \rightarrow X_j \}$ 。

由定义可知, P_c 是一个I-图。

第二步:对 P_c 图进行有效的删除

从定理3中的性质2就可得到一个判断 X, Y 是否条件独立的算法:当给出一个概率分布 $P(x)$ 时,可以通过判断 $I(X, Y \mid Z) = 0$ 来代替 $I(X, Z, Y)$,即当 $I(X, Y \mid Z) = 0$ 时,则有 $I(X, Z, Y)$ 条件独立性成立,从而 P_c 图中的 $X \rightarrow Y$ 和 $Y \rightarrow X$ 边可以删除;否则在给定条件 Z 的情况下, X 和 Y 则互相依赖。

然而在实际计算中并没有一个真正的概率分布 $P(x)$,只是有一个基于样本数据集 D 而计算的一个经验概率分布 $P_D(x)$ 来近似估计 $P(x)$,计算的 $I(X, Y \mid Z)$ 只是基于 $P_D(x)$ 上的 $I_D(X, Y \mid Z)$ 的近似值,所以它的值总是大于0的。为此,可把判断条件独立方法描述为:设 X, Y, Z 为全集 U 上三个不相交的子集,基于样本数据集 D 上概率分布 $P_D(x)$,如果有: $I(X, Y \mid Z) < \varepsilon$,则判定给定 Z , X 与 Y 条件独立;否则给定 Z , X 与 Y 是条件依赖的。其中 ε 为一个阈值,通常取一个很小的正数。

设 U 为属性全集,我们依次对每一对属性 X_i 和 $X_j \in U(i, j = 1, 2, \dots, n), X_i \neq X_j$ 进行计算条件互信息 $I(X_i, X_j \mid U - X_i - X_j)$,给定一个阈值 ε ,当 $I(X_i, X_j \mid U - X_i - X_j) < \varepsilon$ 时,从 P_c 中删除边 $X_i \rightarrow X_j$ 和 $X_j \rightarrow X_i$,这样通过 $O(N^2)$ 次的CI测试, P_c 图中只保留了那些有依赖关系的结点之间的有向边。

由定理2可知,经过这一步删除,在不考虑边的方向的情况下, P_c 图是一个最小I-图。

第三步:确定 P_c 图边的方向

对于任一相互依赖的邻接结点 (X, Y) ,仅存在 $X \rightarrow Y$ 和 $Y \rightarrow X$ 两种情况。Jie Cheng等^[3]提出了一种基于V-结构的边的方向确定方法,但它并不保证能发现所有的有向边。根据打分函数的可分解性^[2],我们提出一种基于 X 和 Y 之间的打分函数。

我们设计一个函数,以度量 X 对 Y (或 Y 对 X)的依赖度,根据概率论中方差的概念,做以下定义:

定义9:设 X, Y 为两个相互关联的结点,且满足假设3,定义:

$f(X \rightarrow Y) = \sum_{X \text{ 的所有取值}} \sum_{Y \text{ 的所有取值}} \sigma^2(p(Y \mid X))$ 的值为结点 Y 对结点 X 的依赖度,其中 $\sigma^2(p(Y \mid X)) = (p(Y \mid X) - \bar{\mu}_Y)^2$, $\bar{\mu}_Y$ 为概率均值。

由定义可知: $f(X \rightarrow Y)$ 的值反应了当变量 X 确定时,变量 Y 的确定度,当 $f(X \rightarrow Y)$ 越大,说明当变量 X 确定时,变量 Y 的确定度就越大,即 Y 对 X 得依赖就越大,反之就越小。

设 X, Y 为两个相互关联的结点,其 Bayesian 网边的方向由以下方法确定:

1) 当 $|f(X \rightarrow Y) - f(Y \rightarrow X)| \leq \varepsilon$ 时,则 $X \rightarrow Y$ 和 $Y \rightarrow X$ 的可能性都存在。其中 ε 为一个小的正数。

2) 当 $|f(X \rightarrow Y) - f(Y \rightarrow X)| > \varepsilon$ 时,当 $f(X \rightarrow Y) > f(Y \rightarrow X)$

时则选取 $X \rightarrow Y$ 弧,删除 $Y \rightarrow X$ 弧;否则选取 $Y \rightarrow X$ 弧,删除 $X \rightarrow Y$ 弧。

由算法可知,在处理过程中可能存在 $X \rightarrow Y$ 和 $Y \rightarrow X$ 两条边都有的情况,这说明在 Bayesian 网中结点 X 和结点 Y 存在着互相依赖,比如:在教学系统中,学生和老师两个结点就是一对互相依赖的结点。在这种情况下,一个系统就存在多种可能性的 Bayesian 网。

第四步:专家评价

利用专家知识对第三步产生的多种可能性的 Bayesian 网进行评价,选择一些最有可能性的结果作为我们的最终目标。

基于信息论的 Bayesian 网构造算法如下:

1) 输入: 样本数据集 D , 结点集 U , 阈值 $\varepsilon_1, \varepsilon_2$

2) FOR I=1 TO N /构造完全有向图 V

FOR J=1 TO N

IF (I≠J) THEN

V = V ∪ { < X_I, X_J > }

ENDIF

END I, J

3) FOR I=1 TO N-1 /根据条件独立性对 V 进行删除

FOR J=I+1 TO N

计算 I(X_I, X_J | U - X_I - X_J)

IF I(X_I, X_J | U - X_I - X_J) < ε₁ THEN

V = V - { < X_I, X_J >, < X_J, X_I > }

ENDIF

END I, J

4) FOR I=1 TO N /根据依赖度确定边的方向

FOR J=1 TO N

IF < X_I, X_J > IN V THEN

计算 F(X_I→X_J) 和 F(X_J→X_I)

IF |F(X_I→X_J) - F(X_J→X_I)| > ε₂ THEN

IF F(X_I→X_J) > F(X_J→X_I) THEN

V = V - { < X_J, X_I > }

ELSE

V = V - { < X_I, X_J > }

ENDIF

ENDIF

ENDIF

END I, J

5) 输出: V

由以上算法可知:整个算法的计算复杂度为 $O(N^2)$ 的 CI 测试。

4 实例分析

此例来自对华盛顿高级中学的 10318 名高年级学生的升学计划的调查,每个学生用下列变量及其相应的状态来描述:

性别(X1):男、女

社会经济状态(X2):低、中下、中上、高

智商(X3):低、中下、中上、高

家长的鼓励(X4):低、高

升学计划(X5):是、否

样本数据:表 1 中的数据表示对于 5 个变量取值的某种组合统计所得到的人数,例如:第一个数据 4 表示对(X1 = 男, X2 = 低, X3 = 低, X4 = 低, X5 = 是)这种组合所统计出的人数,第二个数据 349 表示对(X1 = 男, X2 = 低, X3 = 低, X4 = 低, X5 = 否)这种组合所统计出的人数。变量依次从右到

(下转第 10 页)

套几何命题,用程序实现。在实验阶段,库容量较小语模为简单句型,句型统一。因此,程序全部正确的实现了从自然语言到作图语言的转换。从实验结果来看,以上的构建方法是可行的,达到了预期的效果。

表 1 语料库各类词素数量

类别	数量	占百分比
几何名词	64	0.63
动词	28	0.27
助词	10	0.1

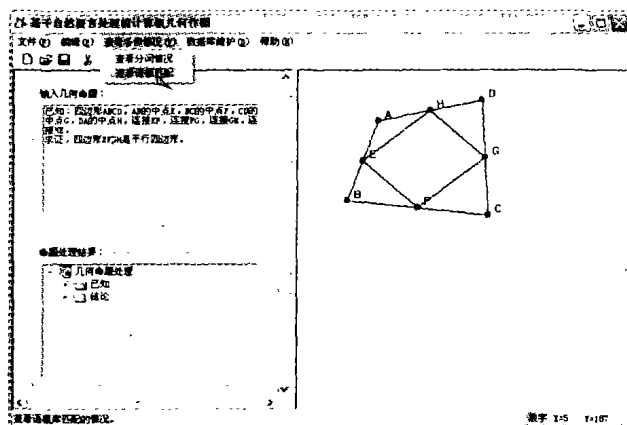


图 3 程序实现界面

9 结语

(上接第 3 页)

左的顺序轮换,状态则按照上面所列各变量状态的顺序进行轮换,依此类推,得到完全统计数据如下:

4, 349, 13, 64, 9, 207, 33, 72, 12, 126, 38, 54, 10, 67, 49, 43, 2, 232, 27, 84, 7, 201, 64, 95, 12, 115, 93, 92, 17, 79, 119, 59, 8, 166, 47, 91, 6, 120, 74, 110, 17, 92, 148, 100, 6, 42, 198, 73, 4, 48, 39, 57, 5, 47, 123, 90, 9, 41, 224, 65, 8, 17, 414, 54, 5, 454, 9, 44, 5, 312, 14, 47, 8, 216, 20, 35, 13, 96, 28, 24, 11, 285, 29, 61, 19, 236, 47, 88, 12, 164, 62, 85, 15, 113, 72, 50, 7, 163, 36, 72, 13, 193, 75, 90, 12, 174, 91, 100, 20, 81, 142, 77, 6, 50, 36, 58, 5, 70, 110, 76, 12, 48, 230, 81, 13, 49, 360, 98

Heckerman 等^[2]用基于统计打分搜索算法得到图 1(a)和(b)两种最有可能的 Bayesian 网结构:

基于我们所给出的算法计算过程如下:

1) 取阈值为 $\varepsilon_1 = 0.007, \varepsilon_2 = 0.001$ 计算结果如图 1(c)。

2) 根据专家知识可知:性别、社会经济状态是不会有父结点的,所以对 $X1 \rightleftharpoons X4$ 和 $X2 \rightleftharpoons X3$ 两种依赖关系可修订为 $X1 \Rightarrow X4$ 和 $X2 \Rightarrow X3$,由此可得 Bayesian 网为如图 1(d)。

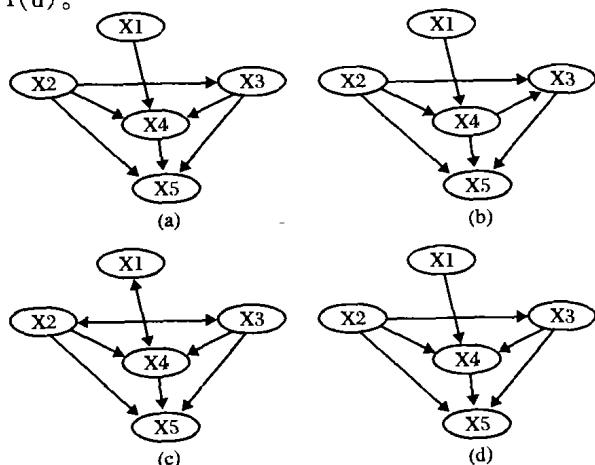


图 1

本文讨论了建立在初等几何命题范围上的自然语言理解系统及其实现,在计算机理解数学语言描述的问题上走出了重要的一步,为今后在更广范围的数学语言的计算机理解提供了借鉴和基础。该系统可作为语音、图形理解的中间件,能广泛的应用于数学教育软件中。

目前,在复合句型和复合结构的问题上缺乏通用对策,且系统所依赖的数据库都是事先生成的,由新词汇、新句型而产生的问题只能通过人工扩充来解决,系统还不具有自学习能力。

参考文献:

- [1] DEAN T, ALLEN J, ALOMONOS Y. *et al.* Artificial Intelligence: Theory and Practice[M]. Publishing House of Electronics Industry, 2002. 489-537.
- [2] 陆亦斌. 自然语言表述初等几何命题的理解[D]. 中国科学院成都计算机应用研究所硕士学位论文, 2000.
- [3] 穗志方, 俞士汶. 汉语单句谓词中心词识别知识的获取及应用[J]. 北京大学学报(自然科学版), 1998, 34(2-3): 220-230.
- [4] 靳光瑾, 陆汝占. 从汉语句子中提取逻辑函子的一种方法[J]. 软件学报, 1998, 9(6).
- [5] 周强, 黄昌宁. 基于局部优先的汉语句法分析方法[J]. 软件学报, 1999, 10(1).

观察以上结果,可发现图 1(d)和图 1(a)是一样的。

5 结语

本文基于信息论中测试信息独立理论,对 Bayesian 网中各结点进行条件独立(CI)测试,以发现各结点的条件依赖关系,并通过计算结点之间的互相依赖度以发现 Bayesian 网边的方向,取消了大多算法要求结点有序的假设,从而构造 Bayesian 结构,算法的计算复杂度只需要进行 $O(N^2)$ 次 CI 测试,经实例证明该算法是行之有效的。但是为了降低搜索空间,本文使用了结点 X 与其父结点的依赖是相互独立的假设,实际上,在有些时候父结点对子结点的影响是联合作用的,特别当结点多,系统复杂的情况下更是如此,所以如何在保证算法效率的同时取消这个假设,是我们今后进一步的研究方向。

参考文献:

- [1] PEARL J. Probabilistic reasoning in intelligent system: Network of plausible inference[M]. San Francisco, CA: Morgan Kaufmann, 1988.
- [2] HECKERMAN D. A tutorial on learning with Bayesian Network [R]. Technical Report MSR-TR-95-06 Microsoft Research, March 1995.
- [3] CHENG J, GREINER R, KELLY J, *et al.* Learning Bayesian networks from data: an information-theory based approach[J]. Artificial Intelligence archive, 2002, 137(1-2): 43-90.
- [4] 刘惟一, 田雯. 数据模型[M]. 北京: 科学出版社, 2001.
- [5] 何盈捷. 基于扩展关系模型的 Bayesian 网的发现[D]. 云南大学硕士研究生学位论文. 云南大学, 2001.
- [6] 姜丹, 钱玉美. 信息论与编码[M]. 北京: 科学出版社, 1992.
- [7] Haykin S. 神经网络原理[M]. 北京: 机械工业出版社, 2004.