

# 信度网推理——方法及问题(上)\*

Inference in Belief Network-Methods and Problems(I)

刘启元 张 聪 沈一栋

(重庆大学计算机科学与工程学院 重庆400044)

**Abstract** Belief network(BN), as a kind of knowledge representation based on the probabilistic theory, has become the main branch in the area of non-deterministic knowledge representation and inference in artificial intelligence(AI) in this decade. BN is being applied in many areas such as industry, military, medical treatment, commerce and so on, and its representative systems involve expert system, voice recognition, fault diagnosis in space crafts, causality mining based on probabilistic syntax, etc. This paper introduces the primary methods and the current problems in inference of BN. After the introduction of the concept, representation and goal of inference of BN, the paper summarizes the basic idea and problems in the principal algorithms of exact inference in BN, which include polytree propagation algorithm, clique tree propagation algorithm, graph reduction algorithm, and combination and optimization algorithm.

**Keywords** Belief network, Probabilistic inference, AI

基于概率知识表达的信度网,已成为人工智能非精确知识表达与推理领域近10几年来研究的热点。目前国外的许多研究机构都对信度网进行了深入的研究。这些研究主要集中在以下三个方面:基于信度网的推理、基于信度网的学习和基于信度网的应用。其中基于信度网的推理一般分为:精确推理(即精确计算概率值)和近似推理(近似计算概率值)两个部分,主要研究高效的推理算法<sup>[1,6]</sup>;基于信度网的学习一般分为参数学习和结构学习两个内容,同时根据样本数据的不同性质每一部分均包括:实例数据完备、实例数据不完备两个方面<sup>[7,6]</sup>;基于信度网的应用,主要包括:基于信度网的知识表达、相应的软件工具开发、基于信度网的实例应用等。目前这些研究都取得了丰硕的成果,正逐步走向实际应用。信度网的提出人 Pearl 教授也于1999年被授予IJCAI 杰出研究成果奖。前微软公司总裁 Bill Gates 在洛杉矶时报上曾说过:微软公司的成功将在于它在信度网方面研究的领先性。在微软的拳头产品:Windows 2000和 Office 系列已经在很多方面融入了信度网,同时微软的一些其它产品如: Pregnancy and Child Care Center 也是基于信度网智能内核开发研制的。目前信度网已成功用于许多其它领域,如:工业上的故障诊断(如美国通用电气公司的 Auxihary

Turbine Diagnosis)、航天故障诊断(如美国航空航天局和 Rockwell 公司联合研制的 Diagnosis of space shuttle propulsion systems)、军事目标自动识别、作战意图自动估计、无人自动驾驶、医学上的病理诊断、商业上的经融市场分析、信息智能检索、基于概率因果关系的数据挖掘、软件产业上的智能帮助系统等。

在文[20][21]中重点介绍了信度网的学习算法,本文将重点介绍信度网的推理算法。在介绍了这类算法的共同特点后,将重点介绍:Polytree Propagation、Clique Tree Propagation、Graph Reduction、Combination and Optimization 方法的基本思想和存在的问题。

## 1 信度网

信度网(Belief Network)<sup>[6,7,11]</sup>又名贝叶斯网络(Bayesian Network),是目前非精确知识表达与推理领域最有效的理论模型,已成为近十几年来研究的热点。该网络采用图形化的方式表达了一个联合概率分布。一个由N个节点 $\{X_1, \dots, X_n\}$ 构成的信度网 $B(G, P)$ ,由两部分构成:

1. 一个具有N个节点的有向无环图:G。该图中的节点为随机变量,节点间的有向边表达了节点间的相关关系。整个图形结构蕴含了如下节点间的条件独立

\* ) 本文受国家自然科学基金、教育部跨世纪优秀人才培养基金以及重庆市科技攻关项目《面向工业应用的智能开发平台及系统研究》的支持,刘启元 博士生,主要研究方向:人工智能技术,信度网推理与学习,因果网推理等。

性:

$$X_i \perp \{X_1, X_2, \dots, X_{i-1}\} \setminus Pa_i | Pa_i$$

(它表示在给定父节点  $Pa_i$  下, 节点  $X_i$  同它的其它上层节点  $\{X_1, X_2, \dots, X_{i-1}\} \setminus Pa_i$  独立)

n. 一个与每个节点相关的条件概率表:  $P(X_i | Pa_i)$ . 它表达了节点  $X_i$  同其父节点的相关关系: 条件概率。

如此, 一个信度网可以表达如下一个联合概率分布<sup>[11,12]</sup>:

$$P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i | X_{i-1}, \dots, X_1) = \prod_{i=1}^n P(X_i | Pa_i)$$

图1所示为一个典型的信度网:

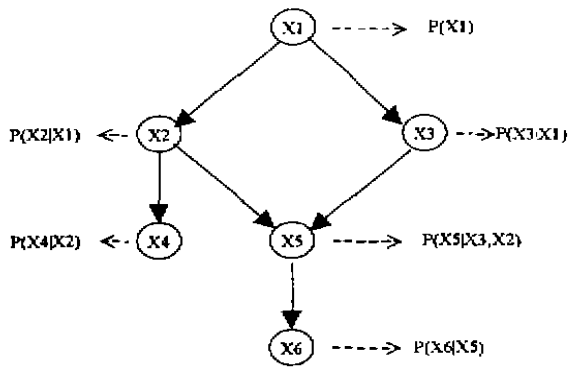


图1 一个典型的信度网结构

该网络由6个节点构成, 蕴含了如下的条件独立性:

$$X_6 \perp \{X_1, X_2, \dots, X_4\} | X_5, X_5 \perp \{X_1, X_4\} | X_2, X_3, X_4 \perp \{X_1, X_3\} | X_2, X_3 \perp \{X_2\} | X_1, X_2 \perp \{X_2\} | X_1$$

表达了如下一个联合概率分布:

$$P(X_1, \dots, X_6) = P(X_6 | X_5) * P(X_5 | X_3, X_2) * P(X_4, X_2) * P(X_3 | X_1) * P(X_2 | X_1) * P(X_1)$$

## 2 概率推理

为了更清楚地表述信度网推理领域的概貌, 本文将从基础的概率推理开始介绍。作为信度网理论基础的概率论, 可以看作是一个组合问题。变量间的联合概率分布是求解所有概率问题的基础。一个由  $n$  个变量所构成的联合概率分布表达了该  $n$  个变量各种组合的概率值。该组合一共包含  $\prod_{i=1}^n \text{Num}(V_i)$  个元素 (其中  $\text{Num}(V_i)$  为变量  $V_i$  的状态数)。为了表述简单, 在本文中未明确指出的情况下, 均假设变量的状态数为2。因此含有  $n$  个变量的联合概率分布一共包含  $2^n$  个概率

值。在概率计算中 一个主要的算子就是边缘化 (marginalizing), 它的计算结果将得到边缘概率分布 (marginal probability) 而其余的概率值 (如: 后验概率) 都可以通过对某些边缘概率进行运算而得到。在边缘概率的计算过程中包含了大量的求和运算, 以消去多余的变量。例如求边缘概率  $P(V_i)$ , 可以计算如下:

$$P(V_i) = \sum_{\{V_j\}} P(V_1, \dots, V_n)$$

其中  $\sum_{\{V_j\}}$  表示对集合  $V$  中去掉变量  $V_i$  后的所有变量求和。如果采用上面的办法直接计算边缘概率, 需要完成不合实际的指数式求和运算量, 为了提高计算速度, 目前研究概率推理算法的主要目的是寻求一种办法以减少该求和量, 其中主要的思想是使得求和计算局部化<sup>[15,16]</sup>。其中 Shannon 和 Shafer 提出了一个基本原理<sup>[15,16]</sup>, 该原理描述了一个计算问题在满足了一些简单的假设下, 如何进行局部计算的问题。它对于一大类问题, 包括: 动态编程、约束满足、概率计算等都有指导作用。目前在概率推理领域, 已设计出多种推理算法如: Polytree Propagation, Clique Tree Propagation, Symbolic Probabilistic Inference (SPI), Bucket Elimination, Graph Reduction 等。虽然经 Cooper 证明<sup>[13]</sup> 在任意结构的信度网上的推理计算是 NP-难的, 但是在实际问题中往往能够根据不同的信度网结构寻找到可以满足要求的推理算法。

正如 Shafer 和 Shenoy 指出的那样, 概率推理方法的共同之处都是首先寻找一种方式对联合概率分布进行参数化, 然后在参数化的基础上寻找局部化的计算过程以加快推理计算速度。而它们之间的不同之处在于每种方法参数化的方式不同或寻求局部计算的途径不同<sup>[15,19]</sup>。在这些参数化的方式中最直接的就是信度网, 它直接利用随机变量间的条件独立性, 将一个联合概率分布直观地表达为一个图形结构和一系列的有条件概率表, 其中条件概率表为该参数化方式的参数。在图1所示的例子中, 一共有6个随机变量。在参数化以前表达一个联合概率分布, 需要给出  $2^6$  即64个参数。而采用信度网的表达方式以后只需要给出  $2+2+2+2+2+2=14$  个参数。实际上对一个联合概率分布的参数化表达了该联合概率分布的内部结构。通过对联合概率分布内部结构的表达不仅提供了一种有效表达联合概率分布的方式, 而且为设计快速推理算法提供了可能性。例如采用直接计算的方式计算  $P(X_5)$  的边缘概率分布, 其计算公式如下:

$$P(X_5) = \sum_{X_1, X_2, X_3, X_4, X_6} P(X_1, X_2, X_3, X_4, X_5, X_6)$$

而采用 SPI 方法, 其计算公式可变换为:

$$P(X_1) = \sum_{X_2} P(X_2 | X_1) * \sum_{X_3} P(X_3 | X_2, X_1) * \sum_{X_4} P(X_4 | X_3, X_2, X_1) + \sum_{X_2} P(X_2 | X_1) * P(X_2 | X_1) + P(X_2 | X_1) + P(X_2)$$

可以看出在2值的情况下,如果采用直接计算一共需要2<sup>3</sup>即32次求和运算,而采用SPI方法则只需要8+8+8+2=26次求和运算,如果在3值情况下,采用直接计算需要3<sup>3</sup>即243次求和运算,而采用SPI方法则只需要27+27+27+3=84次求和运算。通常在信度网的节点数较少的情况下,计算性能的提高感觉不明显,但在一般情况下随着节点数的增多其计算效率的提高将非常显著。

### 3 信度网推理的主要算法

信度网推理主要是利用信度网这一对联合概率分布进行参数化的方式,快速计算待求概率值的过程。目前主要的推理算法通常可以分为如下几类:

- 基于 Polytree Propagation 的方法<sup>[11-12]</sup>
- 基于 Clique Tree Propagation 的方法<sup>[10]</sup>, 如: Junction tree propagation<sup>[4,9-8]</sup>
- 基于 Graph Reduction 的方法
- 基于组合优化问题的求解方法, 如: Symbolic probabilistic Inference<sup>[1,14]</sup>, Bucket Elimination<sup>[5]</sup>

#### 3.1 基于 Polytree Propagation 的方法

基于 Polytree Propagation 的方法是由 Pearl 于 1986 年提出的,该方法的主要思想是直接利用信度网的图形结构,给每一个节点分配一个处理机,每一个处理机利用相邻节点传递来的消息和存储于该处理机内部的条件概率表进行计算,以求得自身的信度即后验概率,并将结果向其余相邻节点传播。在实际计算中,信度网接收到证据以后,证据节点的信度值发生改变,该节点的处理机将这一改变向它的相邻节点传播;相邻节点的处理机接收到传递来的消息后,重新计算自身的信度,然后将结果向自己其余的相邻节点传播,如此继续下去直到证据的影响传递所有的节点为止。该方法中信度网融合、传播的计算公式如下(其推导过程和详细的解释见文[11]):

$$BEL(A_i) = \alpha * \lambda_i(A_i) \dots \lambda_M(A_i) * \sum_{j=1}^K \pi_A(P_{1j}) \dots \pi_A(P_{Nj}) P(A_i | P_{1j}, \dots, P_{Nj})$$

$$\pi_{C_j}(A_i) = \alpha * \prod_{j=1}^M \lambda_j(A_i) * \sum_{P_1^1, \dots, P_1^N} P(A_i | P_1^1, \dots, P_1^N) \pi_A(P_1^1) \dots \pi_A(P_1^N)$$

$$\lambda_A(P_i) = \alpha * \sum_{j=1}^M \prod_{l=1}^M \lambda_l(A_i) * \sum_{P_1^1, \dots, P_1^{j-1}, P_1^{j+1}, \dots, P_1^N} \pi_A(P_1^1) \dots \pi_A(P_1^{j-1}) \pi_A(P_1^{j+1}) \dots \pi_A(P_1^N) P(A_i | P_1^1, \dots, P_1^N)$$

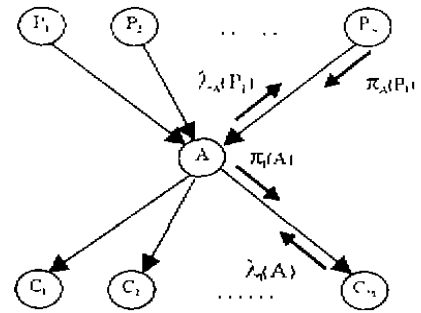


图2 通常情况下的信度值和传递参数值的计算公式

其中:

- $\lambda_i(A_i)$  为子节点  $C_i$  传递给节点  $A_i$  的消息向量,  $\lambda_i(A_i)$  为该消息向量的第  $j$  个元素;
- $\pi_A(P_j)$  为父节点  $P_j$  传递给节点  $A_i$  的消息向量,  $\pi_A(P_j)$  为该消息向量的第  $j$  个元素;
- $\lambda_A(P_j)$  为节点  $A_i$  传递给其父节点  $P_j$  的消息向量,  $\lambda_A(P_j)$  为该消息向量的第  $j$  个元素;
- $\pi_{C_j}(A_i)$  为节点  $A_i$  传递给其子节点  $C_j$  的消息向量,  $\pi_{C_j}(A_i)$  为该消息向量的第  $j$  个元素。

从上可见,采用 Polytree Propagation 算法,每一个节点内部的处理器在运算时都只使用自身的条件概率表和相邻节点传递来的  $\lambda$  和  $\pi$  消息。如果将每一个节点的计算时间平均看作一个单位,则采用该方法的计算时间将正比于证据在传播过程中所经历的最长路径中的节点数,在最坏情况下将正比于信度网网络直径。而每个节点处理机的计算时间将随着该节点父节点的个数的增多呈指数增长。在通常情况下,父节点数不会太多,因此该算法的计算时间复杂度为  $O(W)$  (其中  $W$  为网络直径)。但是该算法仅仅适合于信度网为单连通图情况下的推理。在多连通图的情况下,由于消息传递的双向性,使得消息会在无向环路中循环传播而无法进入稳态,得不到最终结果<sup>[7,11-13]</sup>。为此,许多学者提出了各种弥补的办法,如 Conditioning 方法、Node Aggregation 方法、Star Decomposition 方法……<sup>[11]</sup>。这些方法的主要思想都是对原信度网进行变换,将其由多连通的拓扑结构变换为单连通的拓扑结构,然后再利用 Polytree Propagation 算法进行计算,最后对计算的结果进行处理以还原为待求概率值。其中 Conditioning 方法通过引入额外的条件节点,使得原来多连通的图形结构在条件节点的作用下,满足单连通特性——即在条件节点集合  $C$  中的各节点取值已知的情况下,根据 Pearl 的 D-Separation 判定准则,任意两个节点间的有向通路最多一条,该方法在进行推理时,不仅仅实例化证据节点,而且实例化条件节点集合中的各个节点。对于条件节点集合中各节点的不

同组合都需要进行计算,最后对这些结果加权求和以得到待求概率。它的具体计算公式如下。

$$P(X|Observes) = \alpha \sum_C P(X|Observes \& C) \cdot P(C|Observes)$$

该方法求解的第一个步骤是寻找一个最小组和状态数的条件节点集合,该问题经证明是 NP-难的<sup>[2]</sup>,因此只能希望在多项式时间内寻找一个次优解<sup>[1,7,10]</sup>。Node Aggregation 方法的主要目的是通过对信度网中某些节点进行聚合形成聚集节点,以将一个多连通的图形结构变成一个单连通的图形结构。然后利用单连通图推理算法进行推理,而 Star Decomposition 方法通过引入新的节点将原信度网变成单连通,然后再进行推理(如图3所示)。该方法虽然好,但是能否使用该方法有一个条件,即原信度网的任意三个节点的联合概率分布必须满足:Star-Decomposition 准则。

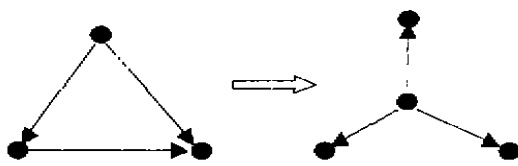


图3 Star Decomposition 方法的基本思想

(未完待续)

#### 参 考 文 献

- 1 Li Zhaoyu, D' Ambrosio B. Efficient Inference in Bayes Networks As A Combinatorial Optimization Problem In *International Journal of Approximate Reasoning* 1994
- 2 Andersen S K, et al. HUGIN-A shell for building Bayesian belief universes for expert systems. In: Proc. of the 11th intl joint conf. on artificial intelligence, Morgan Kaufmann, San Mateo, 1989. 1080~1085
- 3 Becker A, Geiger D. Approximate Algorithms for the Loop Cutset Problem. In: Proc. of 10<sup>th</sup> Uncertainty in Artificial Intelligence, 1994
- 4 Cooper G F. Computational complexity of probabilistic inference using Bayesian [Becker and Geiger 1996] Ann Becker and Dan Geiger, A sufficiently fast algorithm for finding close to optimal junction trees. In: Proc. of 12<sup>th</sup> Uncertainty in Artificial Intelligence 1996
- 5 Dechter R. Bucket Elimination: A unifying framework for probabilistic inference. In: Proc. 12<sup>th</sup> Conference on Uncertainty in Artificial Intelligence, Portland, OR, 1996. 211~219
- 6 Introduction to Bayesian Networks: A Tutorial for the 66th MOR Symposium June 1998. 23~25
- 7 A Tutorial on Learning With Bayesian Networks [Technical Report MSR-TR-95-06]. 1995
- 8 Huang C, Darwiche A. Inference in Belief Networks: A Procedural Guide. *International Journal of Approximate Reasoning*, 1994, 11: 1~158
- 9 Jensen F V, Lauritzen S L, Olesen K G. Bayesian updating in causal probabilistic networks by local computation. *Computational Statistics Quarterly*, 1990, 4: 269~282
- 10 Lauritzen S L, Spiegelhalter D J. Local computations with probabilities on graphical structures and their application to expert systems (with discussion), *J. Roy. Stat. Soc. Ser.* 1988, B50(2): 157~224
- 11 Pearl J. Fusion, propagation, and structuring in belief networks. *Artificial Intelligence*, 1986, 29: 241~288
- 12 Pearl J. *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann, Palo Alto, 1988
- 13 Pearl J. Belief networks revisited, *Artificial Intelligence*, 1993, 59: 49~56
- 14 Shachter R, D' Ambrosio B, DelFavero B. Symbolic probabilistic inference in belief networks. In: proceedings Eighth National Conference on AI, pages 126-131. AAAI, August 1990
- 15 Shafer G, Shenoy P P. Local computation in hypertrees. [Technical Report WP-201]. School of Business, University of Kansas, 1988
- 16 Shenoy P P, Shafer G. Axioms for probability and belief-function propagation. In *Uncertainty in artificial intelligence IV*, Shachter R D, et al eds. North-Holland, Amsterdam, 1990. 169-198.
- 17 Stillman J. On heuristics for finding loop cutsets in multiply connected belief networks. In: Proceedings of the Sixth Conference on Uncertainty in Artificial Intelligence, Cambridge, Massachusetts, 1990. 265~272
- 18 Suermondt H, Cooper G. Probabilistic inference in multiply connected belief networks using loop cutsets, *Int. J. Approx. Reasoning*, 1990, 4: 283~306
- 19 Zhang Nevin Lianwen, Poole D. A simple approach to Bayesian network computation
- 20 邢永康, 沈一桢. 信度网中条件概率表的学习. *计算机科学*, 2000, 27(10)
- 21 邢永康, 沈一桢. 学习信度网的结构. *计算机科学*, 2000, 27(10)