

基于全局消息传播的贝叶斯推理^{*})

刘震 谭良 周明天

(电子科技大学计算机学院卫士通安全联合实验室 成都 610054)

摘要 在贝叶斯网络中,常常需要作不确定概率推理。然而针对一般复杂网络,精确推理算法由于计算复杂度太高而常常被摒弃。针对这一问题,本文提出了一种基于全局传播的 PPJT 近似推理算法。PPJT 算法采用消息传播机制,通过消息的收集与分发过程,可以更新和修正连接树节点的团势并最终生成相容连接树。与另一种常用的近似推理算法即似然权重(Likelihood Weighting)算法的时间性能对比实验显示,采用消息传播机制的 PPJT 算法有效地降低了计算的时间复杂度;同时与似然权重算法的性能对比实验表明,在相对小规模观察样本输入条件下,PPJT 算法能够保证更高的概率推理精度。PPJT 算法为实现一般复杂网络中的概率推理提供了一种新的理论工具。

关键词 概率传播, 贝叶斯网络, 势函数, 消息传递

Bayesian Inference Based on Global Message Propagation

LIU Zhen TAN Liang ZHOU Ming-Tian

(Westone United Lab of College of Computer Science and Engineering, UESTC, Sichuan, Chengdu 610054)

Abstract Uncertain probabilistic inference is often made in Bayesian network. However, for a common complicated network, accurate inference algorithm is always deserted for its unpaid high cost of computing complexity. Aiming at this problem, this paper brings forward a nearly accurate inference algorithm PPJT. New algorithm applies the mechanism of passing message to update the potentials of Join tree's cliques by steps of message collection and message distribution and eventually generates a consistent join tree. Compared with another nearly accurate inference algorithm, namely likelihood weighting algorithm, the time-using performance experimentation shows that PPJT decreases the time complexity efficiently. At the same time, PPJT improves the uncertain inference accuracy. The experimentation for computing accuracy comparison shows that, under relative small samples input, PPJT can ensure much higher accuracy for inference. PPJT provides a new theoretic tool for implementation of probabilistic inference in the common complicated network.

Keywords Probability propagation, Bayesian network, Potential function, Message pass

1 引言

贝叶斯网络也称为信念网络,最早由 Judea Pearl 于 1988 年提出的贝叶斯网络实质上是一种基于概率的不确定性推理网络^[1]。Lauritzen 和 Spiegelhalter 利用影响图理论(Theory of Influence Diagram)对贝叶斯网络加以了深化与发展^[2,3]。通过学习与推理,贝叶斯网络可以利用一些先验的知识对一些现象进行识别、分类和预测。作为一种基于概率的不确定性推理方法,贝叶斯网络在处理不确定信息的智能化系统中得到了重要的应用,已成功地应用于医疗诊断、统计决策、专家系统等领域。

贝叶斯网络推理算法大致可分为精确算法和近似算法两大类。理论上,所有类型的贝叶斯网络都可以用精确算法来进行概率推理。但 Cooper 指出,非单联贝叶斯网络中的精确概率推理是一个 N-P 难题^[5]。对于某一特定拓扑结构的网络,其复杂性取决于节点数。所以,精确算法一般仅用于结构较为简单的单联(Single connected)网络。但对于一般复杂网络,由于不希望计算成为多项式复杂度问题,因此在多数情况下需要采用近似算法。虽然近似算法不能够提供每个节点的

精确概率值,但近似算法的优点是可以较大地简化计算和推理过程。

通过对贝叶斯网络构建连接树是解决一般复杂拓扑结构网络推理的一种有效方法。文[4]采用似然权重推理算法研究这一问题。本文提出的是一种基于全局消息传播的 PPJT 算法。新算法与采用基于似然权重的近似推理算法相比可以将推理计算的时间复杂度有效降低,同时能够在较小规模观察样本条件下,保证一般复杂贝叶斯网络推理的精度需求。

2 PPJT 算法原理

PPJT(Probabilistic Propagation in a Join Tree)作为一种概率推理方法需要用到势函数的理论。下面先简要介绍势的概念。

2.1 势(potentials)的基本运算

定义 1 一组变量 X 上的势定义为一个函数,每个变量对应的值 X 称为它的实例(Instantiation)。势函数可以把实例映射为一个非零实数,用 ϕ_X 表示。势可以构成向量甚至矩阵。

• 边缘化(Marginalization)

^{*})基金项目:电子科学基金(No. 51415010101DZ02)。刘震 博士,主要研究方向为智能安全,不确定推理,人工智能等。谭良 博士,主要研究方向为信息安全,中间件技术。周明天 教授,博导,主要研究方向为网络计算,信息安全。

假设有一组变量 Y , 它的势为 ϕ_Y ; 另外有一组变量 $X, X \subseteq Y$; 则

$$\phi_X = \sum_{Y \setminus X} \phi_Y \quad (1)$$

• 乘法 (Multiplication)

给定两组变量 X 和 Y 以及它们的势 ϕ_X 和 ϕ_Y , 若 $Z = X \cup Y$; 则

$$\phi_Z = \phi_X \phi_Y \quad (2)$$

2.2 贝叶斯网络

贝叶斯网络是一种有向无循环图 (DAG)^[3]。图 1 所示是一个典型的一般复杂贝叶斯网络, 它是非单联的。

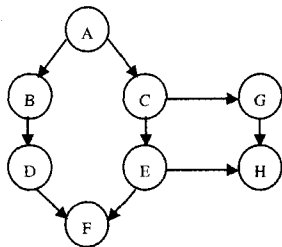


图 1 A belief network

根据 D-separation 原则^[3], 贝叶斯网络的联合概率分布满足

$$P(U) = \prod_{i=1}^n p(V_i | p_{V_i}) \quad (3)$$

V_1, V_2, \dots, V_n 是网络中的变量

图 1 中的联合概率分布为

$$P(U) = p(A) p(B|A) p(C|A) p(D|B) p(E|C) p(F|D, E) p(G|C) p(H|E, G) \quad (4)$$

要对贝叶斯网络作概率推理, 必须先对网络作预处理, 构建连接树 (Join-Tree)。根据文[4], 将该网络道义化 (Moralization), 三角化 (Triangulation) 以后, 采用桶消除算法就可以生成一棵连接树, 如图 2 所示。注意, 因为三角化的方法不唯一, 所以生成的连接树也不唯一。构造最优连接树 (最大团最小的连接树), 也是 NP-Hard 的问题。采用文[5]的方法可以生成近似最优连接树, 如图 2 所示。

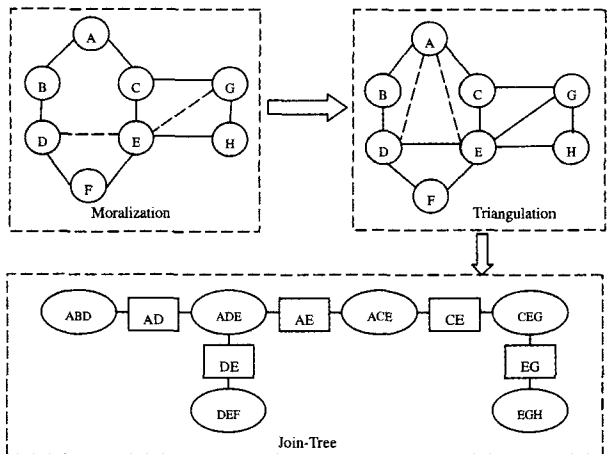


图 2 Preprocessing

对于图 2 中的连接树, 椭圆节点里的一组变量构成一个团 (Clique), 树中任意两个相邻团 X, Y 之间的节点称为团 X, Y 的分割集 (Sepset), 即连接树中的方框节点; 它是团 X, Y 的交集。根据上节对势函数的定义, 团势用 ϕ_X 表示, 分割集势

用 ϕ_S 表示。容易证明贝叶斯网络的联合概率分布可以表示为

$$P(U) = \frac{\prod_i \phi_{X_i}}{\prod_j \phi_{S_j}} \quad (5)$$

2.3 推理原则

由于连接树通常并不相容 (Consistent), 直接作概率推理往往会导致不可信的推理结果^[6], 因此将连接树转换成统计学意义上相容的连接树是 PPJT 算法保证推理精度的关键。采用本文提出的消息传递机制可以对连接树节点势进行重新分配和更新; 全局消息传播的结果可以最终构造出一棵节点势相容的连接树。

2.3.1 初始化 可以量化连接树的势。初始化步骤如下:

- 对每个团和分割集 X , 令 $\phi_X \leftarrow 1$;
- 对网络中的每个变量 V , 令 $\phi_X \leftarrow \phi_X p(V | p_V)$;

以上初始化的过程满足方程

$$\frac{\prod_{i=1}^N \phi_{X_i}}{\prod_{j=1}^Q \phi_{S_j}} = \frac{\prod_{k=1}^Q p(V_k | p_{V_k})}{1} = P(U) \quad (6)$$

N 是团的数量, Q 是网络中节点变量的个数。

2.3.2 单消息传播 (Single Message Pass) 考虑两个相邻的团 X, Y 和它们的分割集 R , 一条消息从 X 传递到要经 Y 过两个步骤:

- 投影 (Projection)

$$\phi_X^{d,d} \leftarrow \phi_R \quad (7)$$

$$\phi_R \leftarrow \sum_{X \setminus R} \phi_X$$

- 吸收 (Absorption)

$$\phi_Y \leftarrow \phi_Y \frac{\phi_R}{\phi_R^{d,d}} \quad (8)$$

注意, 对于任何 R 的实例 r , $\phi_R^{d,d}(r) = 0$ 当且仅当 $\phi_R(r) = 0$ 。如果出现这种情况, 令 $0/0 = 0$ 。式 (7) 和式 (8) 修改了 R 和 Y 的势, 然而仍然保持了方程 (5) 的不变性。因为满足 (9) 式成立。

$$\left(\frac{\prod_i \phi_{X_i}}{\prod_j \phi_{S_j}} \right) \frac{\phi_X^{d,d} \phi_Y}{\phi_R \phi_Y^{d,d}} = \left(\frac{\prod_i \phi_{X_i}}{\prod_j \phi_{S_j}} \right) \frac{\phi_R^{d,d} \phi_Y^{d,d} \phi_R / \phi_R^{d,d}}{\phi_R \phi_Y^{d,d}} = P(U) \quad (9)$$

2.3.3 全局传播算法 全局传播的步骤如下:

- 任意选择一个初始团 X ;
- 调用 COLLECT-MESSAGE(X);
- 调用 DISTRIBUTE-MESSAGE(X);

定义 2 (COLLECT-MESSAGE(X)) 首先标记 X ; 对已被标记的 X 的邻居团, 递归调用 COLLECT-MESSAGE; 如果存在 X 未被标记的邻居团, 从 X 传递一条消息到调用 COLLECT-MESSAGE(X) 的邻居团。

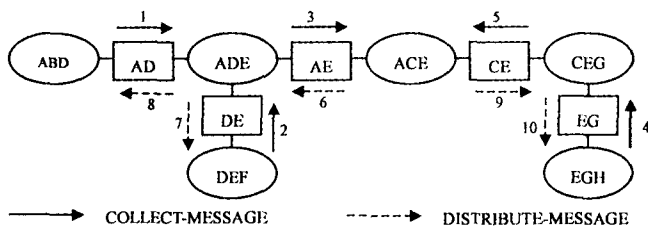


图 3 Global propagation

定义 3(DISTRIBUTE-MESSAGE (X)) 首先标记 X; 如果存在 X 未被标记的邻居团,从 X 传递一条消息到该邻居团;对已被标记的 X 的邻居团,递归调用 DISTRIBUTE-MESSAGE。

图 3 是选择 ACE 作为初始团的连接树全局传播过程示意图。箭头的标号表示一种可能的消息传递顺序。

对于有 n 个团的连接树,全局传播算法要做 2(n-1)次单消息传递。所以,全局传播算法的时间复杂度是线性的 O(n),这与采用似然权重算法的 O(n²)时间复杂度相比有了明显改善。图 4 显示了分别采用 PPJT 算法与似然权重算法的推理时间分布对比,曲线是通过各时间点插值样条拟合得到的光滑线条。从图中可以看出随着网络节点数的增加,采用 PPJT 算法作概率推理的时间明显比变量消除算法所用时间增长缓慢,呈近似线性增长的趋势。但在算法的空间复杂度上,由于全局传播算法要对连接树作两次遍历,这与似然权重算法类似,因此并无优势。

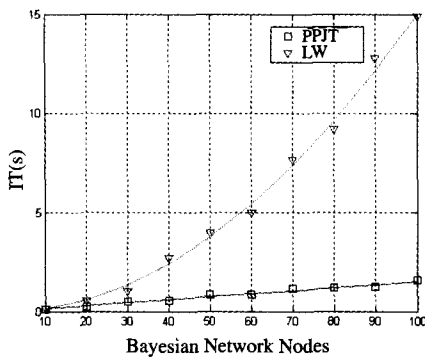


图 4 PPJT 算法与 LW 算法的时间性能对比图

2.3.4 边缘化 通过以上全局传播,我们就建立了一棵相容的连接树,对于任何我们感兴趣的 V,都可以通过以下方法计算 P(V):

- 找到包含 V 的团或者分割集 X;
- 通过边缘化 ϕ_X 计算 $p(V)$ 。

$$p(V) = \sum_{X \setminus \{V\}} \phi_X \quad (10)$$

3 计算实验

以图 3 所示连接树为例,我们利用全局传播算法(PPJT)对网络进行推理计算。首先初始化得到表 1。

表 1 初始化表

Variable	Associate cliques	Potential function
A	A, B, D	$\phi_{A,B,D} = p(A)$
B	A, B, D	$\phi_{A,B,D} = p(A, B)$
D	A, D, E	$\phi_{A,D,E} = p(D B)$
E	A, D, E	$\phi_{A,D,E} = p(D B)p(E C)$
F	D, E, F	$\phi_{D,E,F} = p(F E, D)$
C	A, C, E	$\phi_{A,C,E} = p(C A)$
G	C, E, G	$\phi_{C,E,G} = p(G C)$
H	E, G, H	$\phi_{E,G,H} = p(H E, G)$

经过连接树全局消息传播以后就得到了相容连接树所对应的势函数更新结果如表 2 所示。利用表 2 中的势函数结果,可以边缘化近似求解如图 1 所示的非单联信念网络中的任意节点的分布。

表 2 势函数结果表

Associate cliques	Potential function
A, B, D	$\phi_{A,B,D} = p(A, B)p(D B)$
A, D	$\phi_{A,D} = p(A)p(D B)$
A, D, E	$\phi_{A,D,E} = p(A)p(D B)p(E C)$
D, E	$\phi_{D,E} = p(E C)p(D B)$
D, E, F	$\phi_{D,E,F} = p(E C)p(D B)p(F E, D)$
A, E	$\phi_{A,E} = p(A)p(E C)$
A, C, E	$\phi_{A,C,E} = p(A)p(C E)p(E C)$
C, E	$\phi_{C,E} = p(E C)p(C A)$
C, E, G	$\phi_{C,E,G} = p(G C)p(C A)p(E C)$
E, G	$\phi_{E,G} = p(E C)p(G C)$
E, G, H	$\phi_{E,G,H} = p(G C)p(E C)p(H E, G)$

4 推理精度比较和分析

下面通过计算 KL 散度(KL divergence)^[7],评价概率推理的准确性。KL 值反映的是概率推理分布与真实分布之间的一种“距离”度量,KL 值越小表明近似推理结果与真实分布越接近。

$$KL(Q \| P) = \sum_X Q(X) \log \frac{Q(X)}{P(X)} \quad (11)$$

其中 Q(X)是真实分布, P(X)是通过推理学习获得的分布。

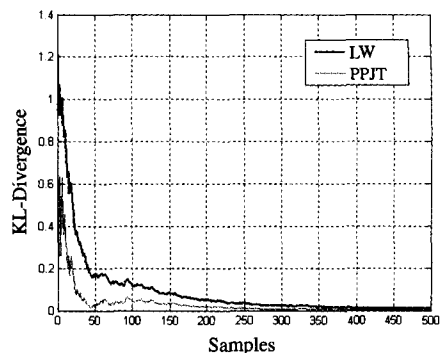


图 5 KL 散度分布图

为了比较 PPJT 算法与似然权重算法的推理精度,本文以图 1 网络中的节点 E 为例,分别采用这两种算法对其分布作概率推理。图 5 是将推理结果作 KL 散度化后得到的对比图。注意,由于在小样本先验的条件下,极大似然方法的置信度太低^[7],因此这里的概率分布先验值的计算均采用了具有共轭分布性质的 Dirichlet 先验作贝叶斯参数估计。从图中可以看出在观察样本个数少于 50 时,两种算法的 KL 散度值都偏高,而且振荡严重。即是说此时两种算法的推理精度在当前数量的样本输入条件下都比较差。但与似然权重算法相比,基于 PPJT 算法的 KL 散度值明显收敛的更快,这是由于全局消息传播得到的相容连接树保证了更高推理精度的结果。观察样本个数超过 50 以后,由两种算法得到的 KL 值均开始逐渐趋于收敛,但始终 PPJT 算法得到的 KL 值较似然权重算法所得到 KL 值更小。所以,在相对小规模观察样本输入条件下,PPJT 算法的推理精度已明显优于似然权重算法。

结束语 基于连接树的全局传播算法(PPJT)利用消息传递机制,通过消息的收集和分发实现连接树团势的更新,并生成了相容的连接树,改善了在非单联贝叶斯网络中作概率推理的时间性能。与似然权重算法相比,时间复杂度从二项式

(下转第 203 页)

点

3. 否则,对于 C_{ijk} 中的每一个对象 P ;

I. 将 $Count_{2,ijk}$ 的值赋给 $Count_p$

II. 对于 $L_2(C_{ijk})$ 中的每一个对象 Q , 计算 PQ 之间的距离 $dist$, 如果 $dist \leq D$, 则 $Count_p$ 加 1, 若 $Count_p > M_{L_2}$, P 不是孤立点, 执行 6(c)(3); 否则, 标记 P 是孤立点。

在算法中, 从 step1 到 step6(c)(2) 都是以单元格为基本单位来判定孤立点, 而不是以像素为单位, 这将大大减少像素间的运算, 从而减少算法的时间复杂度, 提高了边缘检测的效率; step6(c)(3) 是以对象为单位来判断该对象是否为孤立点, 在计算过程中只需计算对象与其所在单元的第二邻居层中对象间的距离, 而不需计算第二邻居层以外的其它对象, 从而也减少了计算时间。

4 实验结果

为了测试算法的有效性, 选择了一幅彩色头像照片进行实验。照片扫描后存储为 BMP 格式的图像文件, 图像的色彩是 24 位, 大小为 150×200 总供包含 30000 个像素。下面对实验结果进行说明。

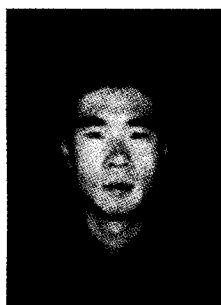


图 1



图 2

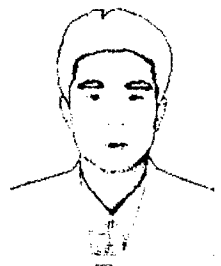


图 3

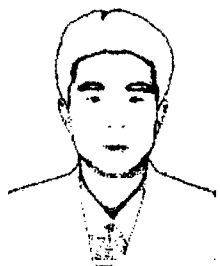


图 4

图 1 是原图像, 图 2、3、4 是当输入的 $D=50, M=70, 100$ 、

150 时, 所得的结果图。由图像可以清楚地看到, 随着 M 的变大, 头像的边缘变得越来越清晰, 此时, 我们得到的孤立点数分别为 734、1266、1703 个像素。因而, 可以通过调节 M 的值, 来得到不同的孤立点数目, 从而也就得到了不同清晰度的边缘图像。同样我们也可以通过调整 D 的值来取得边缘不同清晰度的图片, 下面是取 $M=12, D$ 分别为 23.9、20、17.9, 与此相应的单元格个数分别为 50653、54834、125000, 发现的孤立点分别为 734、1271、1734。在 M 不变的情况下, 随着单元格数目的增加, 所得到的孤立点的数目也相应的增加, 从而使提取的图像边缘也越来越清晰。下图是提取的结果。



图 5

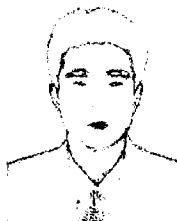


图 6



图 7

对比两种调整方式不难发现, 在所获得的孤立点数目大致相同的条件下, 固定 D 只调整 M 所提取的图像边缘要比固定 M 只调整 D 所提取的清楚, 但是, 这种方式(固定 D , 调整 M)对人面部特征的提取却不够理想, 比较一下, 这两种方式对人嘴的提取, 不难看出, 后一种方式(固定 M 调整 D)能够得到较好的结果, 因而, 在提取的过程中, 可以根据需求的不同, 选择合适的提取方式。

结论 对于人脸的边缘提取问题, 基于单元的孤立点分析算法是一种高效、准确的方法, 它不仅能够准确地提取彩色图像人脸边缘, 而且可以通过调整 D, M 的值来获得不同清晰度的边界, 从而增加了算法的灵活性。试验结果证明了 RGB-OEA 算法的有效性。

参考文献

- 1 Duda R O, Hart P E. Pattern Classification and Scene Analysis. Wiley, New York, 1973
- 2 Kirsch R. Computer Determination of the Constituent structure of Biological Images. Computer and Biomedical Research, 1971, 4 (3): 315~328
- 3 Petrovskiy M I. Outlier Detection Algorithms in Data Mining Systems. Programming and Computer Software, 2003, 29(4): 228~237
- 4 Han Jiawei, Kamber M. Data Mining Concepts and Techniques, 2001
- 5 Knorr E, Ng R, Tucakof V. Distance-Based Outliers: Algorithms and Applications. Very Large Data Bases (VLDB) Journal, 2000
- 6 don; UCL Press, 1996, 196~202
- 7 Lauritzen S L, Spiegelhalter D J. Local computations with probabilities on graphical structures and their application to expert systems [J]. Roy. statist. Soc. Ser B, 1988, 50(11): 17~24
- 8 Lepar V, Shenoy P P. A Comparison of Lauritzen -Spiegelhalter, Hugin and Shenoy -Shafer Architectures for Computing Marginals of Probability Distributions [M]. In G. Cooper and S. Moral, editors, UAI, 1998. 328~337
- 9 Dagum P, Luby M. An optimal approximation algorithm for Bayesian inference [J]. Artificial intelligence, 1997, 93: 1~27
- 10 Shafer G R, Shenoy P P. Local computation in hypertrees [R]. Technical Report 201, School of Business, Uni. Kansas, 1988
- 11 Murphy K. Active learning of causal Bayes net structure [R]. Technical report, Comp. Sci. Div., UC Berkeley, 2001
- 12 Ng A Y, Jordan M I. On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes [C]. In NIPS-14, 2002. 245~250
- 13 Spirtes P, Glymour C, Scheines R. Causation, Prediction and Search [M]. MIT Press, 2000. 78~85

(上接第 168 页)

复杂度降低为线性复杂度; 同时在较小样本条件下与似然权重算法的推理精度对比实验表明, PPJT 算法可以保证更高的不确定推理精度。对于解决如何在一般复杂信念网络中应用概率推理实现可信的统计决策问题, PPJT 算法提供了一种高效可行的解决思路和方法。下一步迫切需要研究的工作内容是如何提高更小样本条件下的推理精度。因为对于诸如风险分析这类问题常常无法提供这样大的样本先验, 而且共轭先验条件也本非总能得到满足。随着针对信念网络的不确定推理方法和不确定学习方法的发展, 针对这类问题的研究必将会受到越来越广泛的关注^[8]。

参考文献

- 1 Jensen F V. An Introduction to Bayesian Networks [M]. Lon-