信度网 条件概率表 学习 贝叶斯统计 (23)

计算机科学2000Vol.27№.10

信度网中条件概率表的学习*)

Learning CPT in Belief Networks

88-92 O212.8

邢永康 沈一栋

（重庆大学计算机科学与工程学院 重庆400044）

Abstract Learning with a belief network is to build its two basic components from data: a network structure and a set of conditional probability tables. It is of crucial importance in the practical application of belief networks, so is one of the current hot research topics. In this paper, we discuss approaches to learn conditional probability tables from a set of data, given a fixed network structure. We will analyze two representative exact algorithms and three widely used approximate algorithms: maximum likelihood estimation, maximum a posterior, EM algorithm, gradient ascent algorithm and Gibbs sampling algorithm, in order for users to choose based on their advantages and shortcomings.

Keywords Belief network, Learning, CPT, Incomplete data, Complete data

一、引言

信度网 B 的学习包括结构 B_s 的学习和条件概率表 B_p 的学习。因果马尔可夫条件原理表明[1]：如果图形 G 是一个随机变量集合 X 的因果图，那么图形 G 也是该随机变量集合的联合概率分布所对应的信度网的结构图。根据这一原理，在实际应用中，可以利用领域专家知识来构造关于这些随机变量的因果图，该图即可作为信度网的结构。但是，信度网的条件概率表通常都比较大（若 X 有 n 个父结点，每个父结点有 M 种取值，则 X 的条件概率表将有 M^n 行），很难让专家来逐项给出。因此，从实际数据中归纳、学习信度网的条件概率表便成了一个很有意义的研究课题。本文讨论在信度网结构给定的情况下，条件概率表的学习方法。

用于信度网学习的学习库可以表示为 $C=\{\vec{C}_1,\vec{C}_2,\cdots,\vec{C}_m\}$，其中第 i 个实例 $\vec{C}_i=[X_1=x_1,X_2=x_2,\cdots,X_n=x_n]$，是对所有随机变量构成的向量 $(X_1,X_2,\cdots,X_n)$ 的一种赋值（包括空值 null）。如果一个实例不含空值，则称该实例是完整的；否则为不完整的。由完整的实例构成的学习库称为完整学习库；否则称为不完整学习库。如图1给出了一个简化的肺部诊断的信度网结构和一个不完整的学习库。

因为信度网是联合概率分布的图形表示，所以信度网的条件概率表学习问题可以归结为统计学中的参数估计问题。对于参数估计问题，统计学中有大量的研究。其方法从基本思想上看，可以分为两大类：一类是基于经典统计学的学习，另一类是基于贝叶斯统计学的学习。

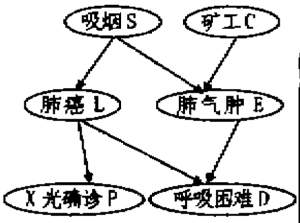
为了便于表达，对条件概率表分层引入如下的表示符号：

$$\vec{\theta}=[\vec{\theta}_1,\vec{\theta}_2,\cdots,\vec{\theta}_n];\vec{\theta}_i=P(X_i\mid parent(X_i))$$

$$\vec{\theta}_i=[\vec{\theta}_{i1},\vec{\theta}_{i2},\cdots,\vec{\theta}_{ir}];\vec{\theta}_{ij}=P(X_i\mid Pa_i^j)$$

$$\vec{\theta}_{ij}=[\theta_{ij1},\theta_{ij2},\cdots,\theta_{ijq_i}];\theta_{ijk}=P(X_i=X_i^k\mid Pa_i^j)$$

其中：Pa_i^j 表示随机变量 X_i 的父结点的取值组合中第 j 个取值组合。用 q_i 表示变量 X_i 的可取值数目，根据概率的归一性质，它们满足等式：$\sum_{k=1}^{q_i}\theta_{ijk}=1$。



No	S	C	L	E	P	D
1	1	1	1	0	1	1
2	0	1	0	1	0	1
3	1	1	1	null	1	1
4	1	1	null	1	1	1
5	1	0	1	0	1	1
6	0	0	0	1	0	1
7	1	0	null	0	1	0
8	0	0	0	null	0	1

图1 关于肺部疾病诊断的信度网络图及一个不完整的学习库

*）国家自然科学基金及教育部跨世纪优秀人才基金资助项目。邢永康 博士生，研究方向：人工智能、知识工程；沈一栋 教授，博士生导师，研究方向：人工智能。

二、基于经典统计学的学习

经典统计学，又称为传统统计学，是基于对概率的频度性理解而建立起来的一系列理论。建立在该理论上的参数估计方法有矩法估计、极大似然估计等。矩法估计要求分布函数已知，而信度网中条件概率表的学习是在不知道分布函数的情况下进行的，所以只能采用极大似然估计方法[illegible]。它建立在极大似然原理基础上，其基本思想是：一个随机实验有若干个可能的结果 $C^1, C^2, \cdots, C^n$，若在一次实验中，结果 C^n 出现，则一般认为实验条件对 C^n 出现有利，也即 C^n 出现的概率应该最大，因此可以将似然函数 $P(C|\theta)$ 取极大值时的参数值 θ 作为对参数的估计值。

1 完整的实例数据

为了研究的方便，一般假设各个参数 $\vec{\theta}_{ij}$ 之间相互独立[1]。对于完整的实例数据 C，假设各个实例数据之间相互独立，根据信度网结构 B_s 所包含的条件独立性关系，可以对参数 $\vec{\theta}$ 的似然函数 $P(C|\vec{\theta})$ 进行如下变换：

$$
\begin{aligned}
L(C:\vec{\theta}) &= \prod_{m}\prod_{i=1}^{n} P(X_i[m]\,|\,Pa_i[m], \vec{\theta}_i) \\
&= \prod_{i=1}^{n}\prod_{m} P(X_i[m]\,|\,Pa_i[m], \vec{\theta}_i) \\
&= \prod_{i=1}^{n}\prod_{j=1}^{q_i}\prod_{m:\,Pa_i[m]=Pa_i^j} P(X_i[m]\,|\,Pa_i^j, \vec{\theta}_{ij}) \\
&= \prod_{i=1}^{n}\prod_{j=1}^{q_i} L(C:\vec{\theta}_{ij})
\end{aligned}
$$

由上式可以看出，似然函数 $L(C:\vec{\theta})$ 的求极值问题就可以分解为对各个因子 $L(C:\vec{\theta}_{ij})$ 的求极值问题。下面计算 $L(C:\vec{\theta}_{ij})$ 的极大值。由于给定参数及父结点后，各个实例相互独立，所以参数 $\vec{\theta}_{ij}$ 的似然函数为：

$$L(C:\vec{\theta}_{ij}) = P(C|\vec{\theta}_{ij}) = \prod_{k}^{r}\theta_{ijk}^{N_{ijk}} \tag{1}$$

上式中，r 表示变量 X_i 的可取值的数目；N_{ijk} 表示在实例数据集合中，变量 X_i 取第 k 个值，父结点集合取第 j 个取值组合的实例的个数，称为充分统计量，因为它们总结了实例数据对似然函数所贡献的信息。

(1)式是一个关于向量变量 $\vec{\theta}_{ij} = [\theta_{ij1}, \theta_{ij2}, \cdots \theta_{ijk}, \cdots]$ 的函数，根据函数极值的求法，该函数的极大值必在它的驻点，即在其导数等于0的点，则：

$$\frac{\partial(P(C|\vec{\theta}_{ij}))}{\partial\vec{\theta}_{ij}} = 0 \Rightarrow \frac{\partial\left(\prod_{k=1}^{r}\theta_{ijk}^{N_{ijk}}\right)}{\partial\vec{\theta}_{ij}} = 0$$

通过计算最终得：

$$\theta_{ijk} = \frac{N_{ijk}}{\sum_{k} N_{ijk}} \tag{2}$$

(2)式就是在完整实例数据下，利用极大似然法对信度网中各个参数的估计公式，利用该公式，可以很容易地计算出给定结构的信度网的条件概率表。

2 不完整的实例数据

当实例数据不完整时，似然函数的计算将变得很复杂，要精确计算其极大值几乎不可能，因此，只能近似地求出似然函数的极大值，并将该点的参数值作为估计值，常用的方法有 EM 算法[illegible]和梯度上升算法[illegible]。

EM(Expectation-Maximization)算法是 Dempster[1997]提出的。利用该算法建立信度网的条件概率表是基于这样一种想法：当实例数据不完整时，无法通过对实例数据的统计来获取充分统计量 N_{ijk}，但是，利用信度网的推理算法，可以对不完整实例中没有观测值的变量进行估计，从而形成一个完整的实例数据集合，再利用公式(2)就可以求出参数 $\vec{\theta}_{ij}$ 的极大似然估计值。

EM 算法：

第一步：给参数 $\vec{\theta}$ 指定一个初始值（一般随机地给定）

第二步：计算每一个充分统计量 N_{ijk} 对于分布 $P(X|\vec{\theta}, C)$ 的期望值 $E_{P(X|\vec{\theta},C)}N_{ijk}$。

$$E_{P(X|\vec{\theta},D)}(N_{ijk}) = \sum_{l=1}^{n} P(X_i^k, Pa_i^j\,|\,\vec{C}_l, \vec{\theta}) \tag{3}$$

利用公式(3)，对实例数据集合中每一个实例 $\vec{C}_l$ 进行考察，可能有以下几种情况：

(1)变量 X_i 及其父结点集合 Pa_i 中的变量在实例 $\vec{C}_l$ 中都有观测值：如果 X_i 及 Pa_i 的观测值 $X_i = X_i^k$ 且 $Pa_i = Pa_i^j$，那么 $P(X_i^k, Pa_i^j|\vec{C}_l, \vec{\theta}) = 1$；否则，$P(X_i^k, Pa_i^j|\vec{C}_l, \vec{\theta}) = 0$。

(2)变量 X_i 及其父结点集合 Pa_i 中的变量，有一些在实例 $\vec{C}_l$ 中没有观测值。由于此时信度网的结构及条件概率表已给定，所以可以利用信度网推理算法，如消息扩散与汇聚算法、关联树算法等，计算出条件概率 $P(X_i^k, Pa_i^j|\vec{C}_l, \vec{\theta})$ 的值。

第三步：利用上面求出的期望充分统计量，计算所有参数 $\vec{\theta}_{ij}$ 的极大似然估计值，获得 $\hat{\vec{\theta}}$：

$$\hat{\theta}_{ijk} = \frac{E_{P(X|\vec{\theta},\vec{C})}(N_{ijk})}{\sum_{k=1}^{q_i} E_{P(X|\vec{\theta},\vec{C})}(N_{ijk})}$$

第四步：比较参数 $\vec{\theta}$ 与 $\hat{\vec{\theta}}$ 的值，如果两者相差很大，则 $\vec{\theta} = \hat{\vec{\theta}}$，转到第二步继续计算；否则，算法结束，此时 $\hat{\vec{\theta}}$ 就是求得的条件概率表。

EM 算法需要利用信度网的推理算法来计算无观测值的变量的期望值，所以选定一个高效的信度网推理算法是优化 EM 算法的关键，如 Lauritzen[1995]采用的关联树算法[6]就减小了 EM 算法的复杂性。

梯度上升算法是一种利用“梯度”来快速计算函数

极大值的数学方法．其执行过程就是沿着函数的表面一步步向函数的极大值逼近的过程。因为函数在某一点的梯度向量的方向，与函数值在该点变化最快的方向一致，所以该算法可以快速获得函数的极大值．似然函数 $P(C|\vec{\theta})$ 是参数 $\vec{\theta}$ 的多变量非线性函数，如果以各个参数 θ_{ijk} 作为坐标，并将函数的值作为一个高度坐标，绘出似然函数的图形，则该函数的图形是一个多维空间中的曲面。在该曲面上任意选定一个点．作为起点。求出该点的梯度向量．然后沿着梯度向量的方向前进一步，到达函数表面上的一个新点，继续沿着该点的梯度向量前进，反复进行．最终将到达该函数的一个极大值点。

梯度上升算法

第一步：初始化。给参数 $\vec{\theta}$ 随机地赋值，并指定步长 α。

第二步：根据下式．计算似然函数在点 $\vec{\theta}$ 的梯度向量 $\Delta\vec{\theta}$。

$$\frac{\partial \ln P(C|\vec{\theta})}{\partial\theta_{ijk}}=\sum_{l=1}^{n}\frac{P(X_i^k,Pa_i^j|\vec{C}_l,\vec{\theta})}{\theta_{ijk}}$$

该公式是 Russel[1995]证明[5]的，其中的 $P(X_i^k,Pa_i^j|\vec{C}_l,\vec{\theta})$ 可以利用信度网的推理算法（如消息扩散算法、关联树算法）来计算。

第三步：如果该点的梯度向量为0，则算法结束，此时 $\vec{\theta}$ 为求得的条件概率表；如果梯度向量不等于0，则对参数 $\vec{\theta}$ 进行修改．获得新的参数 $\vec{\theta'}=\vec{\theta}+\alpha*\Delta\vec{\theta}$。并进行归一化处理．使参数 $\vec{\theta'}$ 符合条件概率表的约束条件。转第二步继续进行。

需要指出的是．当实例数据不完整时．似然函数有多个极大值。如果以各个参数 θ_{ijk} 作为坐标．并将函数的值作为一个高度坐标．绘出似然函数的图形，则该函数的图形是多维空间中具有多个"凸"点的曲面。因此．EM 算法和梯度上升算法都存在局部最大问题。为了解决这个问题，一般是采用多次循环法。每获得一个极大值后、对参数做些修改．利用梯度上升法重新计算。多次反复．最后取极大值最大的一次计算结果。

三、基于贝叶斯统计学的学习

贝叶斯统计学是统计学中的一个分支．其奠基性工作由贝叶斯完成[7]。贝叶斯统计学与传统统计学的本质区别在于对概率的理解。不同于经典统计学对概率的频度性理解，贝叶斯统计学派认为概率是人们对事物发生可能性的一种合理置信度，具有主观性。建立在该理论上的参数估计问题其基本思想为：给定一个含有未知参数 θ 的分布 $P(X|\theta)$ 以及一个完整的实例数据集合 C，贝叶斯统计学派认为参数 θ 是一个随机变量（区别于经典统计学）．它具有一个先验分布 $P(\theta)$．代表了人们在进行学习以前所知道的关于参数 θ 的信息．要根据以往的知识来估计，如果没有任何知识可参考，可以采用贝叶斯假设．即认为 $P(\theta)$ 是一个均匀分布，θ 在它的取值范围内取到各个值的机会均等。然后通过在实例数据集合 C 上的学习．人们所掌握的参数 θ 的信息发生了变化．表示为 $P(\theta|C)$．称为参数 θ 的后验概率。所以．贝叶斯参数学习的主要任务就是计算参数变量的后验概率 $P(\theta|C)$．并以此作为参数估计的依据。一旦后验分布获得了．则有两种方法估计参数的值：一是用后验分布最大时参数 θ 的值，作为参数的估计值，称为最大后验分布估计（Maximize A Posterior）；另一种是采用后验分布的数学期望作为对参数的估计值．称为条件期望估计。

1. 完整的实例数据

条件期望估计又称为贝叶斯预测法[8]，是指将参数 $\vec{\theta}$ 的后验分布的数学期望作为参数的估计值，即：$\hat{\vec{\theta}}=E_{P(\theta|C)}(\vec{\theta})=\int\vec{\theta}P(\vec{\theta}|C)d\vec{\theta}$，该方法在贝叶斯统计中被大量地采用，因为它充分体现了贝叶斯统计思想。

为了简化计算．一般假设各个参数向量 $\vec{\theta}_{ij}$ 相互独立[13]．称为参数独立性假设。当实例数据集合 C 完整时，可以证明此时参数 $\vec{\theta}_{ij}$ 的后验分布也相互独立[14]。因此可以独立计算各个参数 $\vec{\theta}_{ij}$ 的后验分布 $P(\vec{\theta}_{ij}|C)$，并利用该后验分布分别进行参数估计。

为了计算后验分布 $P(\vec{\theta}_{ij}|C)$．按照贝叶斯统计的思想，必须首先为参数 $\vec{\theta}$ 指定一个先验分布。先验分布的选取有多种方法．如前面的贝叶斯假设，另外还有最大熵原则、杰弗莱原则等。一般最常用的是共轭分布法。当选用共轭分布时．参数的后验概率与先验概率具有相同的形式，可以极大地简化计算。在离散型信度网的学习中，一般将每个分布 $P(X_i|Pa_i^j)$ 看作是多态分布。对于多态分布．其共轭分布是狄尔赫特分布，该分布的定义为：

$$Dir(\vec{\theta}|\alpha_1,\alpha_2,\cdots,\alpha_r)=\frac{\Gamma(\alpha)}{\prod_{k=1}^{r}\Gamma(\alpha_k)}\prod_{k=1}^{r}\theta_k^{\alpha_k-1}$$

其中，α_k 称为超级参数，满足：$\alpha_k>0$；$\sum_{k=1}^{r}\theta_k=1$；$\alpha=\sum_{k=1}^{r}\alpha_k$；$\Gamma(\alpha)$ 函数的定义为：$\Gamma(\alpha)=\int_0^{\infty}t^{\alpha-1}e^{-t}dt$；参数 θ_k 的数学期望为：$E_{Dir(\theta)}(\theta_k)=\int\theta_k Dir(\vec{\theta}|\alpha_1,\alpha_2,\cdots,\alpha_r)d\vec{\theta}=\alpha_k/\sum_{k=1}^{r}\alpha_k$

这里选用狄尔赫特分布作为 $\vec{\theta}_{ij}$ 先验分布后，则其后验分布为：

$$P(\vec{\theta}_{ij}|C)=P(C|\vec{\theta}_{ij})P(\vec{\theta}_{ij})$$

$$= \prod_{k=1}^{r} \theta_{ijk}^{N_{ijk}} \cdot \frac{\Gamma(\alpha)}{\prod_{k=1}^{r} \Gamma(\alpha_k)} \prod_{k=1}^{r} \theta_{ijk}^{\alpha_k - 1}$$

$$= \frac{\Gamma(\alpha)}{\prod_{k=1}^{r} \Gamma(\alpha_k)} \prod_{k=1}^{r} \theta_{ijk}^{N_{ijk}+\alpha_k - 1}$$

可见，后验分布仍然是一个狄尔赫特分布，所以其数学期望可以很容易地求出，并作为对参数的估计，

$$\hat{\theta}_{ijk} = E_{P(\vec{\theta}_{ij}|C)}(\theta_{ijk}) = \frac{\alpha_k + N_{ijk}}{\sum_{k=1}^{r}(\alpha_k + N_{ijk})}$$

上式就是采用条件期望法学习信度网的条件概率表的公式。利用该公式可以依次求出条件概率表中的每一项的值。公式中的参数 α_i 代表专家的知识，在实际计算中，可以由专家采用等价抽样大小的估计方法来估计。特殊情况下，可以采用贝叶斯假设，即假设变量取各个值的概率都相等。此时，

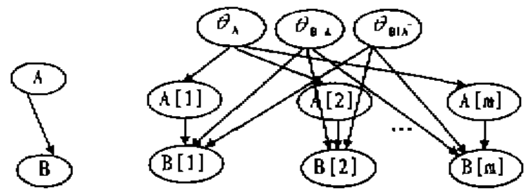
$\alpha_k = 1$ 且 $\alpha = \sum_{k=1}^{r} \alpha_k$。

2. 不完整的实例数据

当实例数据不完整时，采用以上的假设，参数的后验分布已不再是狄尔赫特分布，所以要计算后验分布的数学期望将非常困难，一般采用近似计算——最大后验分布估计法。基本想法是：参数的后验分布是相对于实例数据集合的参数的概率，该值越大，说明实例数据对该参数的支持越大。因此可以将后验分布最大时参数 θ 的值作为参数的估计值。根据贝叶斯公式，可将参数 $\vec{\theta}$ 的后验分布表示为：

$$P(\vec{\theta}|C) = P(C|\vec{\theta})P(\vec{\theta})/P(C) = \alpha P(C|\vec{\theta})P(\vec{\theta})$$

如果采用贝叶斯假设，即认为参数 $\vec{\theta}$ 为均匀分布，由上式可见此时的后验分布的最大点与似然函数的最大点相同。因此，可以采用前面所讲的 EM 算法、梯度上升法来进行最大后验估计。



(a) 简单信度网　　(b) 含有 m 个实例的学习图

(a)简单信度网　(b)含有 m 个实例的学习图

图2　有 m 个实例的学习图

假设给定信度网的结构及 m 个实例，要计算参数 $\vec{\theta}$ 的后验分布。因为每一个实例对应信度网的一个赋值状态，所以可以将这些实例全部用图形表示出来，并将每个参数作为一个结点，从而可以构造出一个扩展的信度网——"学习图"。如图2(b)就是图2(a)所示的信度网有 m 个实例的学习图。给定结构的信度网的条件概率表的学习，就是要求出参数 $\vec{\theta}$ 对于给定的实例数据的后验分布 $P(\vec{\theta}|C)$。这一思想表现在学习图上，就是计算给定证据（实例数据集合 C）下参数变量 $\vec{\theta}$ 的分布，因此，信度网的条件概率表的学习问题就转换成了信度网的推理问题。显然学习图的规模很大，很难采用精确推理算法进行计算，一般采用信度网上的近似推理算法。这里采用最常用的 Gibbs 抽样算法[9-10]。

Gibbs 抽样算法

第一步：初始化：置循环次数 Count＝0；并对实例中没有观测值的变量随机地给定一个初始值，形成完整的实例数据集合 C_c。给每一个参数 $\vec{\theta}_{ij}$ 指定初始值。这里假设参数 $\vec{\theta}_{ij}$ 是一个 Dirichlet 分布，由专家采用等价抽样大小的估计方法来指定超级参数 α_{ijk} 的值。

第二步：任意选取一个没有观测值的变量 X_{il}（实例 $\vec{C}_l$ 中的变量 X_i），计算在给定其它变量的值时 X_{il} 的条件概率。即：

$$P(X_{il}^{k}|C_c \backslash X_{il}) = \frac{P(X_{il}^{k}, C_c \backslash X_{il})}{P(C_c \backslash X_{il})} = \frac{P(X_{il}^{k}, C_c \backslash X_{il})}{\sum_{k=1}^{q_i} P(X_{il}^{k}, C_c \backslash X_{il})}$$

其中，$C_c \backslash X_{il}$ 表示完整实例数据集合 C_c 中去掉变量 X_{il} 的值。由于上式中的每一项 $P(X_{il}^{k}, C_c \backslash X_{il})$ 都表示完整的实例数据集合（X_{il}^{k}&&$C_c \backslash X_{il}$）的概率，因此可以利用下式来计算。

$$P(C_c) = \prod_{i=1}^{n} \prod_{j=1}^{p_i} \frac{\Gamma(\alpha_{ij})}{\Gamma(\alpha_{ij} + N_{ij})} \prod_{k=1}^{q_i} \frac{\Gamma(\alpha_{ijk} + N_{ijk})}{\Gamma(\alpha_{ijk})}$$

该式由 Cooper 和 Herskovits［1992］给出[12]。其中，α_{ij}，α_{ijk} 是超级参数；N_{ij}，N_{ijk} 是多态分布的充分统计量。$\Gamma(\alpha)$ 是一个伽玛函数。

第三步：根据概率 $P(X_{il}|C_c \backslash X_{il})$ 确定变量 X_{il} 状态值。

第四步：对实例数据中所有没有观测值的变量重复第二步与第三步，从而获得一个完整的实例数据集合 C_c。

第五步：由于 C_c 是完整的，因此可以采用前面的数学期望估计法，求出新的参数 $\vec{\theta}'$。

第六步：$\vec{\theta} = \vec{\theta}'$，转到第二步继续计算，并将循环次数 Count 加1。

第七步：经过多次重复计算，最后以参数 $\vec{\theta}$ 的 Count 次平均值作为学习所得的结果。

结束语　从实例数据中学习信度网的条件概率表是信度网走向实际应用必须解决的一个问题。本文探讨了两类学习信度网条件概率表的方法：基于经典统计学的学习与基于贝叶斯统计学的学习。前者具有理论成熟，计算简单的优点，但它只利用了实例数据集合所提供的信息，无法加入专家的知识，所以对实例数据

的依赖性大；而基于贝叶斯统计学学习有机地综合了这两类信息，因此对实例数据集合的依赖性较低，其学习结果更精确。但贝叶斯统计学中的一些理论仍然存在许多争议，如：将参数 θ 看成随机变量是否妥当？参数的先验分布是否存在？在计算中如何选取？等。

在实际应用中，用于学习的实例数据往往是不完整的，这种情况下一般采用近似算法进行学习。常用的近似学习算法有：Gibbs 抽样算法、EM 算法和梯度上升算法。从算法的时间复杂度来看，Gibbs 抽样算法的效率是最低的，当选用相同的推理算法时，EM 算法和梯度上升算法的计算复杂性相同。另外，Gibbs 抽样算法只能用来对信度网的条件概率表进行批量更新，而 EM 算法和梯度上升算法既可以用于对条件概率表的批量更新，又可以用于条件概率表的顺序更新。

参 考 文 献

1 Spirtes P, et al Causation, Prediction, and Search. Spring-Verlag, New York 1993
2 Fisher R A. Theory of Statistical Estimation. In, Proc. of the Cambridge Philosophical Society, 1925, 22: 700～715
3 Dawid A P, Lauritzen S L Hyper Markov laws in the statistical analysis of decomposable graphical models Annals of Statistics, 1993, 21(3)
4 Dempster A, et al. Maximum likelihood form incomplete data via the EM algorithm. Journal of the Royal statistical Society, 1977, B39 1～38
5 Russel S, et al. Local learning in probabilistic network with hidden variables. In: Proc. of the 14-th IJCAI, Montreal, Canada, Morgan Kaufmann, 1995. 1146～1152
6 Lauritzen S The EM algorithm for graphical association models with missing data Computational Statistical and Data Analyses, 1995, 19 191～201
7 Bayes T R An essay towards solving a problem in the doctrine of chance Phil. Trans sic, London, PP53, 370, 1763
8 Heckerman D. A Tutorial on Learning With Bayesian Networks [Technical Report MSR-TR-95-06]. 1995
9 Hrycej T. Gibbs Sampling in Bayesian Networks Artificial Intelligence, 1990, 46: 351～363
10 York J. Use of the Gibbs sampling in expert systems. Artificial Intelligence, 1992, 56, 115～130
11 Heckerman D, Geiger D. Learning Bayesian Networks: the combination of knowledge and statistical data: [Microsoft Technical Report MSR-TR-94-09]. 1994
12 Cooper G, Herskovits E A Bayesian method for induction of probabilistic netwoeds from data Machine Learning, 1992(9): 309～347
13 Spiegelhalter D, Lauritaen S. Sequential updating of conditional probabilities on directed graphical structures. Networks, 1990, 20, 579～605
14 Thiesson B. Accelerated quantification of bayesian networks with incomplete data. In: Proc of First Intl conf. on Knowledge Discovery and Data Mining, Montreal, QU, Morgan Kaufmann, 1995. 306～311

(上接第105页)

在平面上相交的系数依次逐帧排序。

4. 编码

将经 zig-zag 后的512个系数做为一个单位进行熵编码：DC(直流)分量减去前一个单位的 DC 分量，对差值进行 Huffman 编码；AC(交流)分量先进行 RLE 编码，再进行 Huffman 编码。

三、实验结果

我们对 Miss America, Chaire 人像等序列(灰度)分别进行了实验，每帧块进行2-D DCT 变换，帧方向只对每帧次对角线及其左上部进行1-D DCT 变换，右下部不变换并且采用前述量化矩阵进行量比，最后编码。

从实验结果可以看到，对于 Miss America, Claire 等图像序列，由于每帧帧内相关性强，且各帧变化区域小，每个帧块经过2-D DCT 变换后，次对角线右下部的变换系数经量化后，大部分都为零，并且在允许一定失真但保证图像质量的情况下，这些量化后的系数基本都是零。因此采用部分3-D DCT 算法及量化策略可以得到很高的压缩比及很好的解码恢复图像，同时由于采用部分3D-DCT 算法减少了近七分之一的 DCT 运算量，从而提高了编解码效率。

参 考 文 献

1 余英林编著. 图像处理与模式识别. 华南理工大学出版社
2 Westwater R, Furht B Real-Time Vidio Compression Techniques and Algorithm. U. S. A. Kluwer Academis Publishers. July 1997
3 N. 阿罕麦德, K R. 罗. 数字信号处理中的正交变换. 北京 人民邮电出版社, 1979