文章编号:1000 - 582x(2001)01 - 0078 - 06

基于互信息和测度学习信度网结构*

邢永康，沈一栋
（重庆大学 计算机科学与工程学院，重庆 400044）

摘　要：交叉熵是对一个分布与其近似分布的接近程度的度量。在许多关于信度网结构的学习文献中，都将交叉熵作为检验算法学习效果的一个指标。笔者直接从交叉熵最优这一指标出发，在分析已有测度的基础上，提出了一个新的测度——互信息和测度，并证明了该测度的可分解性质。最后，给出了利用互信息和测度进行信度网结构学习的两种启发式搜索算法。

关键词：信度网；信度网结构学习；互信息；结构学习测度

中图分类号：TP 181　　**文献标识码**：A

1 信度网结构学习简介

信度网[1]的学习包括结构学习及条件概率表学习，其中较难的是信度网结构学习。用于信度网结构学习的数据是实际工作中积累的某一领域的历史资料，一般存储在特殊的数据库之中，称为学习数据库。在这里将其通称为实例数据，并以集合的形式描述为：

实例数据集合 $C = \{C_1, C_2, \cdots, C_n\}$ 其中第 i 个实例 C_i 表示该领域随机变量构成的向量的一个赋值，其形式为：$C_i = [X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n]$。

如图 1 是关于肺部疾病诊断的一个实例数据集合，它含有 8 个实例。

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

图 1　一个实例数据集合

由 n 个节点构成的所有有向无环图都可能作为信度网结构，如图 2 就是对图 1 实例数据的一个可能的信度网结构。信度网结构学习的主要任务就是从这些可能结构中找出一个最适合于实例数据的结构[2]。一般方法是首先根据学习任务定义（或选择）一个测度，用来衡量一个结构对实例数据的适合程度。然后计算每一个可能结构的测度，最后通过比较从中选出测度最优的结构。在实际计算中，由于可能结构的数目非常大（由 n 个结点构成的所有有向无环图的数目是结点数目 n 的指数），计算所有结构的测度几乎不可能，所以常采用启发式搜索算法，在结构空间中进行搜索，比较来获得测度最优的结构。

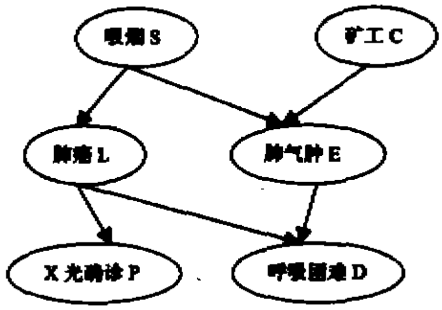

图 2　一个可能的信度网结构

已有许多关于信度网结构学习的文献，提出了一些测度，比较常用的有两个：一是基于贝叶斯统计的 BDe 测度（Bayesian Dirichlet By Likelihood Equivalence）[2]。二是基于编码压缩理论的最小描述长度测度（Minimal Description Length）[3]。由于不同的测度

* 收稿日期：2000-05-17
基金项目：国家自然科学基金项目（69883009）及教育部跨世纪优秀人才基金项目
作者简介：邢永康（1971-），男，陕西人，重庆大学博士生。主要从事人工智能研究。

是根据不同的应用要求而提出的，采用一个测度学习所得的最优结构只能保证在该测度上最优，对于另外的应用要求，利用该测度学习所得的结构不一定是最优的。因此，必须定义一种综合衡量指标，来衡量采用各种测度学习所得结构的最优性。基于信度网的概率本质（信度网是联合概率分布的图形表示），所以公认的的衡量指标是交叉熵。笔者试图从交叉熵出发，基于信息论理论，定义一个新的信度网结构学习测度——互信息和测度。

2　综合衡量指标——交叉熵

交叉熵（Crossing Entropy）的概念[4]最先由 Kullback 与 Leibler 提出，在数字信号处理领域中被广泛地应用。

定义 1　基于随机变量集合 $X = \{X_1, X_2, \cdots, X_n\}$ 的两个分布函数 $Q(\cdot)$ 和 $P(\cdot)$，分布 $Q(\cdot)$ 对于分布 $P(\cdot)$ 的交叉熵为：

$$\mathrm{CE}(P(\cdot), Q(\cdot)) =$$

$$\sum_{[X_1, X_2, \cdots, X_n]} P(X_1, X_2, \cdots, X_n] \log \frac{P(X_1, X_2, \cdots, X_n)}{Q(X_1, X_2, \cdots, X_n)}$$

为了便于表达，在以后的推导中，用 Σ 表示 $\sum\limits_{[X_1, X_2, \cdots, X_n]}$。交叉熵表示两个分布函数所包含的信息的差异程度。两个分布函数差异越大，交叉熵的值就越大；两个分布函数的差异越小，交叉熵的值就越小；当两个分布相等时，交叉熵的值等于 0。

可以认为用于学习的实例数据集合 $C = \{C_1, C_2, \cdots, C_n\}$ 是一个分布函数 $P(X_1, X_2, \cdots, X_n)$ 的随机抽样，一般称该分布为实际分布或标准分布。假设采用一种测度函数，通过对实例数据集合 $C = \{C_1, C_2, \cdots, C_n\}$ 的学习，获得了一个信度网结构及条件概率表，该信度网所表示的联合概率分布为 $Q(X_1, X_2, \cdots, X_n)$，称为学习分布或标准分布的近似分布。结构学习的目标就是使 $Q(X_1, X_2, \cdots, X_n)$ 等于标准分布 $P(X_1, X_2, \cdots, X_n)$。然而由于信息不全以及计算的复杂性等各种原因的限制，学习所得的分布只能尽可能的接近标准分布，而不能完全等于标准分布。两者之间的差异程度可以用交叉熵表示。交叉熵越小，表示该结构越接近于真实结构。正是基于这一点，在许多关于信度网结构学习的文献中，一般都采用交叉熵来衡量结构学习算法的学习效果。既然将交叉熵作为学习效果的衡量指标，那么，能不能直接用交叉熵作为测度，来学习信度网的结构呢？可以设想，如果可以计算出所有可能结构的交叉熵，通过相互比较，其中交叉熵最小的结构就是最优的信度网结构。

要计算出一个信度网结构的交叉熵，还需要确定该结构所对应的条件概率表。

引理　两个分布 P、Q，当 $P(X \mid Y) = Q(X \mid Y)$ 时，$\sum\limits_{X, Y} P(X, Y) \log Q(X \mid Y)$ 的值最大。

证明

$$\max\Big(\sum_{X, Y} P(X, Y) \log Q(X \mid Y) =$$

$$\sum_X \sum_i P(X, Y = Y^i) \log Q(X \mid Y = Y^i) =$$

$$\sum_X \sum_j P(X \mid Y = Y^i) P(Y = Y^i) \log Q(X \mid Y = Y^i) =$$

$$\sum_i P(Y = Y^i) \sum_X P(X \mid Y = Y^i) \log Q(X \mid Y = Y^i)$$

表达式 $\sum\limits_X P(X \mid Y = Y^i) \log Q(X \mid Y = Y^i)$ 是随机变量 X 的函数，由哈夫曼编码理论可知不等式 $\sum\limits_X P(X) \log Q(X) \leqslant \sum\limits_X P(X) \log P(X)$ 始终成立。所以当 $Q(X \mid Y = Y^i) = P(X \mid Y = Y^i)$ 时，表达式 $\sum\limits_X P(X \mid Y = Y^i) \log Q(X \mid Y = Y^i)$ 取最大值。由于对随机变量 Y 的所有取值，上面的情况都成立，所以，当 $P(X \mid Y) = Q(X \mid Y)$ 时，$\sum\limits_{X, Y} P(X, Y) \log Q(X \mid Y)$ 的值最大。

定理 1　设给定一个信度网结构 B_S，该结构表示的分布 $Q_{B_S}(\cdot)$ 是标准分布 $P(\cdot)$ 的近似分布，要使分布 $Q_{B_S}(\cdot)$ 对于标准分布 $P(\cdot)$ 的交叉熵 $CE(P(\cdot), Q_{B_S}(\cdot))$ 最小，则必然存在等式：

$$Q(X_i \mid \mathrm{Parent}(X_i)) = P(X_i \mid \mathrm{Parent}(X_i))$$

证明：

$$\min(CE(P(\cdot), Q(\cdot)) \Rightarrow \min\left(\Sigma P(\cdot) \log \frac{P(\cdot)}{Q(\cdot)}\right) \Rightarrow$$

$$\min(\Sigma P(\cdot) \log P(\cdot) - \Sigma P(\cdot) \log Q(\cdot)) \Rightarrow$$

$$\max(\Sigma P(\cdot) \log Q(\cdot))$$

分布 $Q(\cdot)$ 可以按照信度网结构 B_S 表示为乘积的形式：

$$\max(\Sigma P(\cdot) \log Q(\cdot)) =$$

$$\max(\Sigma P(\cdot) \log \prod_{i=1}^{n} Q(X_i \mid \mathrm{Parent}(X_i))) =$$

$$\max(\Sigma \sum_{i=1}^{n} P(\cdot) \log Q(X_i \mid \mathrm{Parent}(X_i))) =$$

$$\max(\Sigma \sum_{i=1}^{n} P(X_i \,\&\, \mathrm{Parent}(X_i)) \log Q(X_i \mid \mathrm{Parent}(X_i))) =$$

$$\max(\sum_{i}^{n} \Sigma P(X_i \,\&\, \mathrm{Parent}(X_i)) \log Q(X_i \mid \mathrm{Parent}(X_i)))$$

对每一个 i 的表达式 $\Sigma P(X_i \ \& \ \mathrm{Parent}(X_i))\log Q(X_i \mid \mathrm{Parent}(X_i))$,根据引理可知,当 $Q(X_i \mid \mathrm{Parent}(X_i))$ 时,表达式的值最大。从而其求和的值也最大。所以,当 $Q(X_i \mid \mathrm{Parent}(X_i)) = P(X_i \mid \mathrm{Parent}(X_i))$ 时,交叉熵最小。

根据定理 1,当信度网结构确定之后,满足最小交叉熵的条件概率表也随之确定,它与标准分布满足定理 1 所表示的等式。因此,一个给定拓扑结构 B_S 的最小交叉熵为:

$$\min(\mathrm{CE}(P(\cdot),Q_{B_S}(\cdot))) = - \sum \sum_{i=1}^{n} P(X_i \ \& \ \mathrm{Parent}(X_i)) \cdot \log Q(X_i \mid \mathrm{Parent}(X_i))) \tag{1}$$

在该公式中,将一个拓扑结构的交叉熵的计算归结为了标准分布的计算。为了进一步计算,引进实例数据完整性假设。

实例数据完整性假设[2] 假设实例数据集合中每一个实例中不包含空值(Null)

当实例数据完整时,概率的值可以通过对实例数据的统计求出。如变量 X_1 等于 x_1 条件下变量 X_2 等于 x_2 的条件概率值可这样计算:

$$P(X_2 = x_2 \mid X_1 = x_1) = \frac{N_{X_2=x_2 \& X_1=x_1}}{N_{X_1=x_1}}$$

上式中,$N_{X_1=x_1}$ 表示实例数据集合中,变量 $X_1 = x_1$ 的实例数目。可以更一般地定义 N_{ijk},它表示实例数据集合中,变量 X_i 取第 k 值,且变量的父结点集合取第 j 个值的实例的数目。对公式(1)进一步推导:

$$\min(\mathrm{CE}P(\cdot),Q_{B_S}(\cdot))) = - \sum \sum_{i=1}^{n} P(X_i \ \& \ \mathrm{Parent}(X_i)) \cdot \log Q(X_i \mid \mathrm{Parent}(X_i))) = - \sum \sum_{i=1}^{n} \frac{N_{ijk}}{N} \cdot \log \frac{N_{ijk}}{N_{ij}} \tag{2}$$

(2)即为实例数据完整时,一个结构的最小交叉熵计算公式。利用(2)可以很方便地计算给定结构的最小交叉熵,从而可以对多个可能的结构进行比较,取交叉熵最小结构作为所求得的信度网结构。

可以看出,作为算法学习效果的综合指标——交叉熵,可以直接作为测度,用来学习信度网的结构。然而,由于交叉熵没有可分解性质,所以在实际的应用中,要对每个结构分别进行计算,通过比较来选取最优的结构。当变量的数目较多时,可能的结构数目将非常庞大,对每个结构分别计算交叉熵,几乎是不可能的。

3 可分解测度——互信息和

定义 2[2] 结构的测度是可分解的,是指该测度可以依据该结构分解为多项乘积的形式,其中的每一项是一个节点与其父节点的函数,称为本地测度。

例如前面提到的 BDe 测度、最小描述长度测度等就具有可分解性。

为什么要使测度具有可分解性呢?这是因为大多数的搜索学习算法是利用对一个初始的拓扑结构进行连续的修改来产生新的拓扑结构。如对于一个拓扑结构,可以按照规则对它进行修改(任意选取一对节点,如果它们之间有边相连,则可以删除这条边或者将该有向边反向;如果它们之间没有边相连,可以任意方向添加一条边),从而产生新的拓扑结构。而所产生的新的拓扑结构与旧的拓扑结构大部分相同,只有部分结构发生了变化。如果测度具有上述的可分解性,这时不直接求结构的测度,而是求结构的测度的对数。根据测度可分解性定义及对数的性质,则测度的对数就等于各个本地测度的对数和。这样,并不需要对该结构进行重新计算,只需要求出发生变化的本地测度的对数就可以了。从而极大地简化了搜索算法的复杂性。从这个分析也可以看出,测度的可分解性,是为了将计算限制在局部,所以将一个测度分解为本地测度的乘积或者和,都可以达到简化计算的目的。只不过当分解为乘积的形式时,计算的是测度的对数。因此,可以对上面的测度的可分解性定义进行扩展。

定义 3 一个测度是可分解的,是指该测度可以依据信度网结构分解为多项乘积或者和的形式,其中的每一项是一个节点与其父节点的函数,称为本地测度。

测度的可分解性可以简化学习算法的复杂性。前面给出的最小交叉熵,尽管可以用来进行结构的学习,但由于其明显没有可分解性,所以无法直接作为结构学习测度,必须寻找新的可分解测度。

定义 4 两个随机变量 X_1 和 X_2 的互信息为:

$$I(X_1;X_2) = H(X_1) - H(X_1 \mid X_2) = \sum_{X_1} \sum_{X_2} P(X_1,X_2) \log \frac{P(X_1,X_2)}{P(X_1)P(X_2)} \tag{3}$$

其中,$H(X_1)$ 表示随机变量 X_1 的先验熵;$H(X_1 \mid X_2)$ 表示在变量 X_2 给定后,变量 X_1 的后验熵。互信息表示了两个变量之间的制约关系。即当变量 X_2 的值确定后,会使变量 X_1 的不确定性减少(熵减少),互信息就表示了这种熵的减少量。

定理 2　给定一个信度网的拓扑结构 B_S，当其交叉熵最小时，则各个节点与其父节点的互信息和最大。表示为：

$$\min(\mathrm{CE}P(\cdot),Q_{B_S}(\cdot))\Leftrightarrow$$
$$\max(\Sigma I(X_i \mid \mathrm{Parent}(X_i)))$$

证明

$$\min(\mathrm{CE}P(\cdot),Q_{B_S}(\cdot))=\min\left(\Sigma P(\cdot)\log\frac{P(\cdot)}{Q_{B_S}(\cdot)}\right)=$$
$$\min(\Sigma P(\cdot)\log P(\cdot)-\Sigma P(\cdot)\log Q_{B_S}(\cdot))=$$
$$\min(\Sigma P(\cdot)\log P(\cdot)-$$
$$\Sigma P(\cdot)\log\prod_{i=1}^{n}Q_{B_S}(X_i\mid \mathrm{Parent}(X_i)))$$

根据定理 1，且 $\Sigma P(\cdot)\log P(\cdot)$ 与结构 B_S 无关，可以去掉，从而可推出：

$$\Rightarrow\max(\Sigma P(\cdot)\log\prod_{i=1}^{n}P(X_i\mid \mathrm{Parent}(X_i)))=$$
$$\max(\sum\sum_{i=1}^{n}P(\cdot)\log P(X_i\mid \mathrm{Parent}(X_i)))=$$
$$\max\left(\sum\sum_{i=1}^{n}P(\cdot)\log\frac{P(X_i\ \&\ \mathrm{Parent}(X_i))}{P(\mathrm{Parent}(X_i))}\right)=$$
$$\max\left(\sum\sum_{i=1}^{n}P(\cdot)\log\frac{P(X_i\ \&\ \mathrm{Parent}(X_i))}{P(\mathrm{Parent}(X_i))P(X_i)}*P(X_i)\right)=$$
$$\max\left(\sum\sum_{i=1}^{n}P(\cdot)\log\frac{P(X_i\ \&\ \mathrm{Parent}(X_i))}{P(\mathrm{Parent}(X_i))P(X_i)}+\right.$$
$$\left.\sum\sum_{i=1}^{n}P(\cdot)\log P(X_i)\right)=$$
$$\max\left(\sum\sum_{i=1}^{n}P(\cdot)\log\frac{P(X_i\ \&\ \mathrm{Parent}(X_i))}{P(\mathrm{Parent}(X_i))P(X_i)}-\right.$$
$$\left.\sum\sum_{i=1}^{n}P(\cdot)\log\frac{1}{P(X_i)}\right)=$$
$$\max\left(\sum_{i=1}^{n}I(X_i;\mathrm{Parent}(X_i))-\sum_{i=1}^{n}H(X_i)\right)$$

由于 $\sum_{i=1}^{n}H(X_i)$ 与拓扑结构 B_S 无关，可以去掉，则：

$$\Rightarrow\max\left(\sum_{i=1}^{n}I(X_i;\mathrm{Parent}(X_i))\right)$$

该定理表明，可以由交叉熵最小推导出互信息和最大，并且根据定义 3 可以看到，各节点与其父节点的互信息之和明显具有可分解性，所以可以使用互信息和作为测度。

为了简化计算，仍然采用用实例数据完整性假设。当实例数据完整时，与前面计算交叉熵相同，可以很容易地推出，一个信度网结构 B_S 的互信息和测度（Sum of Mutual Information）等于：

$$\mathrm{SMI}(B_S:C)=\sum_{i=1}^{n}I(X_i;\mathrm{Parent}(X_i))=$$
$$\sum_{i=1}^{n}\sum_{k}\sum_{j}\frac{N_{ijk}}{N}\log\frac{N_{ijk}*N}{N_{ik}*N_{ij}}\tag{4}$$

其中，k 表示变量 X_i 的第 i 个值；j 表示 X_i 的父结点集合的第 j 个赋值。

4　启发式搜索算法

信度网的结构学习问题是一个 NP 问题，所以在实际计算中，并不是对所有的结构分别计算测度，通过比较取最优，而是采用启发式搜索算法，按照测度在可能的拓扑结构空间中进行搜索来获取信度网结构。最简单的启发式搜索算法是“瞎子爬山法”[5]。该算法的执行过程很象一个瞎子爬山的过程，由于看不见山峰在那里，他只能用手杖测量四周的地形，以便决定下一步向那里走，他认为山峰应该是在最陡峭的方向。所以他每一步都朝着地形最陡峭的方向前进，在这种指导下，期望最终到达山顶。下面给出利用互信息和测度的“瞎子爬山算法”，该算法可以简单地移植到其它的测度。

爬山算法：

第一步：初始化。任意确定一个信度网结构（可以随机地选取一个特殊的结构，或者依靠专家的知识，建立一个结构），并求出该结构的互信息和。

第二步：对该结构进行修改（如添加一条有向边；删除一条有向边；改变一条有向边的方向等），限制条件是修改后所产生的新结构不能含有有向环。用集合 E 表示对当前结构所有可能的修改构成的集合。对 E 中的每一个修改 e，求出修改后结构互信息和的变化量 $\Delta(e)$。

第三步：如果所有的 $\Delta(e)$ 都小于 0，则停止，此时的信度网结构就为所求；如果不是所有的 $\Delta(e)$ 都小于 0，则采用 $\Delta(e)$ 最大的修改 e 修改当前结构，并计算出结构的互信息和。转第二步，继续。

在第二步求解 $\Delta(e)$ 时，利用测度函数的可分解性，可以减少计算量。例如，假设修改 e 只是给节点 X_1 添加了一条指向它的有向边，此时只有节点 X_1 的互信息发生了变化，其它的节点的互信息都没有变化，因此只要求出节点 X_1 与它的父节点的互信息，就可以计算出 $\Delta(e)$，并不需要对其它的节点进行计算。

爬山算法存在的问题是局部最优问题，即无法保证求得的结果总是全局最优。这可以形象地理解为：当瞎子到了一个小山丘，此时周围的地势都低，瞎子就以为到了山顶而停止前进。解决这种问题的一个方法是

所谓的"重复爬山法"、即求出一个最优的结构之后,再任意地改变当前的结构,并以该结构作为初始结构、利用最陡爬山法继续学习。如此反复数次、则最后可能获得全局最优的结构。

另外一个可以解决局部最优问题的算法是"模拟退火法"。它是 Metropolis 等在 1953 年提出的。在冶金技术上,如果把材料加热到很高的温度、然后让它慢慢冷却,则容易得到晶体结构比较完整,错位和缺陷较少的加工件。这是因为当温度很高时,材料中的分子和原子具有较高的能量,能够较自由的移动,从而跳过那些比较浅的局部最小。Hopfield 曾将这种算法应用于神经网络学习中,用来克服局部最优问题。这里,我们将该算法用于信度网结构学习,简单描述如下:

模拟退火算法:

第一步:初始化。任意确定一个信度网结构(可以随机地选取一个特殊的结构,或者依靠专家的知识,建立一个结构),并确定一个较高的温度值 T_0。另外置循环变量 $i=0$。

第二步:可以对该结构进行修改(如添加一条有向边;删除一条有向边;改变一条有向边的方向等),限制条件是修改后所产生的新的结构不能含有有向环。用集合 E 表示对当前结构所有可能的修改构成的集合。随机地从集合 E 选取一个对当前结构可能的修改 e,求出以下表达式的值:

$$p=\exp\left(\frac{\Delta(e)}{T_0}\right)$$

第三步:如果 $p>0$,则采用修改 e,修改当前的结构;如果 $p<1$,则以概率 p 采用修改 e,修改当前的结构。

第四步:重复第二步和第三步 a 次。如果在 a 次重复中没有修改结构,则停止算法、此时的结构即为所求;否则,循环次数加 1,即 $i=i+1$。如果 $i>\gamma$,则停止算法,此时的结构即为所求。如果 $i<\gamma$、按照降温步长 $\beta(0<\beta<1)$ 降低温度 T_0、即:$T_0=T_0\times\beta$,然后转入第二步,继续进行。

该算法的核心在第二步和第三步,分析表达式 $p=\exp\left(\frac{\Delta(e)}{T_0}\right)$ 可以看出,当随机选取的修改 e 引起的互信息和的变化量 $\Delta(e)$ 大于 0 时,p 的值必然大于 1、根据该算法第三步,修改 e 被采用,这一点保证了算法向互信息和增大的方向前进;当 $\Delta(e)$ 小于 0 时,p 的值必然小于 1,根据该算法第三步、修改 e 被采用的概率为 p,即此时的算法可能向互信息和减小的方向前进、特别当温度值 T_0 很高时,这种可能性增大,当温度值 T_0 降低时,这种可能性减少。由此可见,该算法在计算过程中,并不是一直朝着测度最大的方向前进、有时会沿着测度降低的方向前进,这种"进中有退"的计算避免了陷于局部最优点,从而解决了局部最优问题。

5 结语

基于信息论的信度网结构测度 —— 互信息和测度是根据交叉熵最小直接推倒出来的、所以、利用该测度学习所得的结构,其交叉熵也必然最小。最小交叉熵作为算法学习结果的综合衡量指标是大家公认的、所以利用互信息和测度学习所得的结构必然也是最优的。这一点优于其它的测度,因为别的测度是基于某一个测度的最优,并不能保证对所有的应用目的最优。并且,互信息和测度具有可分解性,它可以直接使用在启发式搜索算法中,加快信度网结构学习的速度、

互信息和测度 SMI 与其它测度的关系可以从以下方面分析:

1) 相对于 BDe 测度,互信息和测度计算比较简单。在 BDe 测度的计算中,需要估计每个结构的先验概率以及它的先验参数;而在互信息和测度中,并不需要计算参数的先验值,所以计算比较简单。

2) MDL 测度定义为:

$$MDL(G:C)=DL_{Stru}(G:C)+N^{*}\sum_{i}I(X_i;Pa_i)$$

第一项表示结构的复杂度;N 表示实例数据个数。因此、如果不考虑结构的复杂度、则 MDL 测度等价于互信息和测度。

需要指出的是,本文中互信息的计算、采用了实例数据完整性假设。当实例数据不完整时、互信息的计算就比较复杂。直观地解决方法是采用一些近似计算方法来补全实例数据,如 Gibbs 抽样法等。另外,如何利用测度函数的特点,进一步加速搜索算法的速度、如何利用专家的知识,来减少搜索的结构空间等,这些还有待遇进一步的研究。

参考文献:

[1] JUDEA PEARL. Fusion, Propagation and Structuring in Belief Networks[J]. Artificial Intelligence, 1986, 29: 241-288

[2] HECKERMAN D, GEIGER D, CHICKERING M. Learning Bayesian networks: The combination of knowledge and statistical data[J]. Machine Learning, 1995, 20: 197-243

[3] RISSANEN. Stochastic Complexity in Statistical Inquiry [J]. World Scientific Publishing Company, River Edge NJ 1989.

[4] KULLBACK S, LEIBLER R. Information and Sufficiency [J]. Ann Math Statistics, 1951, 22: 79-86
[5] 陆汝钤. 人工智能(下)[M]. 北京：科学出版社, 1996

Learning the BN Structure Based on the Sum of Mutual Information

XING Yong-kang, SHEN Yi-dong
(College of Computer Science and Engineering, Chongqing University, Chongqing 400044, China)

Abstract: The Crossing Entropy is defined to scale the similar level of two probability distribution. In many papers on learning BN structure, the Crossing Entropy was used as an indicator of measuring the learning accuracy of an algorithm. The known scoring metrics for learning BN structure is analyzed in this paper, then a new scoring metrics -Sum of Mutual Information is proposed based on the information theory. At last, two algorithm for learning BN structure by SIM is represented.

Key words: belief networks; learning the BN structure; mutual information; scoring metrics

(责任编辑　吕赛英)

※ ※

(上接 69 页)

Multicriterion Optimization for Projection Data in Computerized Tomography

LI Hong[1], LI Ning[2], WANG Xu-dong[2], LI Shi-guang[2], ZHENG Er-xing[2]
(1. CQTV, Chongqing 400039, China; 2. Chongqing University, Chongqing 400044, China)

Abstract: It is an effective way for improving quality of reconstructed images that projection data are optimized in the field of Computerized Tomography (CT). For lowering the fuzziness on projection data caused by lots of factors such as noise and getting the unigue optimal reconstructed image from noisy projection data further more, in this paper, the authors develop a new model named Multicriterion Optimization Model (MOP) for optimizing projection data. The model is structured based on the theories of fuzzy mathematics and decision-making, via the method of deducing projection fuzzy exponent function and square error fuzzy exponent function. The experiments for validating the model have been carried out on personal computer (PC) At first, we make simulation collecting in the images presented and add Gauss Noise into the projection data obtained by simulation collecting, Then, complete the image reconstruction from noisy projection data by the solution without optimization and another solution with multicriterion optimization. Finally, compare and analysis the different results about images reconstructed by two different solutions. The experiment results indicate that the MOP in this paper has better consistency with the theory and practice as well as obvious advantage of antinoise ability.

Key words: computerized tomography; noise; projection data; fuzzy exponent function; multicriterion optimization model

(责任编辑　李胜春)