计算机科学2000Vol. 27№. 10

学习信度网的结构*

Learning Network Structure From Data

邢永康 沈一栋
(重庆大学计算机科学与工程学院 重庆400044)

Abstract A general approach to learn a network structure is to heuristically search the space of network structures for the one that best fits a given data set. The key to the search is a score function which evaluates different network structures. In this paper, we give a detailed introduction to two representative score functions: BDe, MDL. We also discuss two widely used learning algorithms that apply those score functions to direct their search: hill climbing and simulated annealing. Finally, We briefly sketch an algorithm, SEM, that can learn a network structure from incomplete data.

Keywords Belief network, Learning structure, BDe, MDL, SEM

一、等价的信度网结构

学习信度网的结构，就是通过分析实例数据库，建立能够表达实例数据所包含信息的信度网的结构。任何一个由所有的结点(变量)构成的有向无环图都可能作为信度网的结构。如对图1(d)所示的关于吸烟、性别及肺癌的实例数据库，其三种可能的信度网结构如图1中(a)(b)(c)所示。

序号	吸烟(Sm)	性别(S)	肺癌(L)
1	T	M	T
2	F	M	F
3	T	W	F
4	F	W	F
5	T	W	T
	⋮		

d 关于吸烟、性别、肺癌的学习数据库

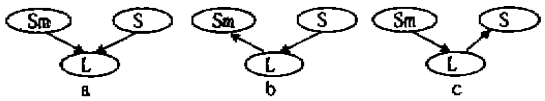

图1 三种可能的信度网结构

根据这三个结构所包含的条件独立性，可以将这三个变量的联合概率分布分别表示为：

$$P_a(Sm,S,L)=P(Sm)P(S)P(L|Sm,S)$$

$$P_b(Sm,S,L)=P(S)P(L|S)P(Sm|L)$$

$$P_c(Sm,S,L)=P(Sm)P(L|Sm)P(S|L)$$

反复应用贝叶斯公式对图b对应的联合概率分布进行变换：

$$P_b(Sm,S,L)=P(S)P(L|S)P(Sm|L)$$

$$=P(S|L)P(L|Sm)P(Sm)=P_c(Sm,S,L)$$

可见，图b与图c所表示的结构对应相同的联合概率分布。信度网可以看成是联合概率分布的图形表示，在这种意义上图b和图c应该是等价的。

定义1[1] 如果两个信度网结构对应相同的联合概率分布，即两个结构中包含了相同的条件独立性，那么这两个信度网结构是等价的。

如果认为信度网结构中的有向边不仅表示变量之间的条件依赖关系，而且表示变量之间的因果关系，这样的信度网称为因果型信度网；否则称为非因果型信度网。显然，若把图b和图c处理成因果型信度网，则它们是不等价的。尽管两者表示了相同的联合概率分布，但却包含着不同的因果关系：图b中，肺癌是吸烟的直接原因，而图c中，吸烟是肺癌的直接原因。限于篇幅，本文只讨论非因果型信度网的学习。对于因果型信度网的学习，可以参见文[2]。因此，非因果型信度网的结构学习就是学习一类等价的结构。

二、信度网结构学习的测度

对于由n个变量构成的信度网来说，n个结点构成的所有有向无环图都可能作为信度网的结构。Robinson[1997]给出了以下递推计算公式，可以求出

*)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。

n 个变量构成的信度网结构的数目：

$$f(n)=\sum_{i=1}^{n}(-1)^{i+1}\frac{n!}{(n-i)!i!}2^{i(n-i)}f(n-i)$$

上述公式表明，当变量的数目增加时，可能的结构数目将以变量数目的指数增加，因此，可能结构的数目将非常庞大，学习信度网的结构，就是要通过分析实例数据集合，从这样大量的结构中选出最适合这些实例数据的网络结构。一般做法是首先定义一个关于结构的测度(Score)，再分别计算出各个可能结构的测度值，最后从中选取测度值最优的结构作为信度网的结构。常用的测度有两个：基于贝叶斯统计的 BDe 测度和基于编码理论的 MDL 测度。

1. 基于贝叶斯统计的 BDe 测度

在 Cooper 及 Buntine 研究的基础上，Heckerman[2]提出了 BDe 测度，该测度以贝叶斯统计学作为理论基础。一个结构 G 的 BDe 测度定义为该结构相对于给定实例数据集合 C 的后验分布。利用 BDe 测度进行信度网结构学习就是寻找该后验分布最大的结构。即：

$$Score_{BDe}(G:C)=P(G|C)\propto P(C|G)P(G)$$

可见，要计算一个结构 G 的 BDe 测度，必须分别计算出实例数据的似然分布 $P(C|G)$和结构的先验分布 $P(G)$。为了简化计算，可以引入以下假设：

1) 多态分布假设　假设根据结构 G 中包含的条件独立关系，可以将实例数据分解为多态分布的集合。每一个多态分布是变量 X_i 对于其父结点集合 Pa_i 的第 j 个取值组合的条件分布，即 $P(X_i|Pa_i^j)$。

根据该假设，结构 G 对应的条件概率表 $\vec{\theta}_G$ 可以表示为以下的形式。其中，每一个 $\vec{\theta}_{ij}$ 是多态分布 $P(X_i|Pa_i^j)$的参数。

$\vec{\theta}_G=\{\vec{\theta}_{ij}|i=1,\cdots,n,j=1,\cdots,q_i\}$，$q_i$ 表示 Pa_i 的所有取值组合的总数目。

2) Dirchlet 假设　假设参数 $\vec{\theta}_{ij}$ 的先验分布是 Dirchlet 分布。

3) 参数独立性假设　假设各个参数 $\vec{\theta}_{ij}$ 的先验分布之间相互独立，即

$$P(\vec{\theta}_G)=\prod_i\prod_j P(\vec{\theta}_{ij})$$

当实例数据 C 完整时，根据以上三条假设，可以计算出结构 G 的 BDe 测度为：

$$Score_{BDe}(G:C)=P(G|C)$$

$$=P(G)\prod_i\prod_j\frac{\Gamma(\sum_k N_{ijk})}{\Gamma(\sum_k N_{ijk}+N_{ij})}\prod_k\frac{\Gamma(N_{ijk}+N_{ijk})}{\Gamma(N_{ijk})}$$

其中，N_{ijk}表示实例数据集合 C 的充分统计量，可以通过对实例数据的统计求得，$P(G)$是结构 G 的先验分布，N_{ijk}是每个参数 θ_{ijk} 分布的超级参数。这两者在对一个结构进行测度计算之前必须预先给出。当可能的结构空间很大时，要人工对每一个结构进行估计明显是不可能的，Cooper 和 Herskovits[1992]采用了简单的估计方法，指定所有的 $N_{ijk}=1$，并假设所有可能的结构出现的概率相同，即 P(G)是一个均匀分布，由此对应的测度称为 K2测度。该测度计算较为简单，但这两个假设过于武断，Heckerman 引入以下两个假设：

4) 参数模块性假设　对于两个不同的信度网结构 G，G′，如果一个变量 X 在这两个结构中具有相同的父结点集合，那么该变量的条件概率表在两个结构中应该相同。

5) 似然等价性假设　对于两个等价的信度网结构 G，G′，它们的参数也具有等价性，即：

$$P(\vec{\theta}_G|G)=P(\vec{\theta}_G|G')$$

在此基础上，对于非因果型信度网的学习，只要由专家预先估计出一个关于变量的联合概率分布 $P_{p0}(X)$及等价样本大小 N'，就可以计算出所有其它结构的超级参数 N_{ijk}。

$$N_{ijk}=N'*P_{p0}(X_i^k,Pa_i^j|G_c)$$

由于直接估计联合概率表比较抽象，Heckerman[4]进一步指出，可以由专家根据自己的知识，预先构造出一个完整的信度网——先验信度网(prior network)，来表示专家对这些变量的联合概率分布的估计。利用先验信度网，一个结构的先验分布定义为：$P(G)=ck^{\delta}$，其中，c 为一个常数，$0<k\leqslant 1$，δ 表示当前结构与先验信度网的结构不相同的边的数目。即：

$$\delta=\sum_{i=1}^{n}\delta_i.$$

2. 基于编码理论的 MDL 测度函数

最小描述长度(MDL)原理是 Rissanen[5]在研究通用编码时提出的。其基本原理是：对于一组给定的实例数据 C，如果要对其进行保存，为了节省存储空间，一般采用某种模型对其进行编码压缩，然后再保存压缩后的数据。同时，为了以后完全恢复这些实例数据，也要求对所使用的模型进行保存，所以需要保存的数据长度等于这些实例数据进行编码压缩后的长度加上保存模型所需的数据长度，将该长度称为总的描述长度。最小描述长度原理就是要求选择总描述长度最小的模型。

如果将信度网作为对实例数据进行压缩编码的模型，MDL 原理就可用于信度网的学习。一个信度网可以表示出实例数据中每个实例的分布，利用该分布，如果采用 Hufferman 编码(出现频率最大的数据，其编码长度最小)，就可以达到压缩保存的目的。当然，所使用

的信度网也必须保存。所以总的描述长度就是压缩后的数据长度加上保存该信度网结构所需的数据长度。根据 MDL 原理，信度网学习的任务就是要找到一个信度网结构，用该信度网对实例数据进行压缩保存，总的描述长度最小。因此，一个结构的 MDL 测度由以下两部分构成：

1）描述信度网的数据长度　保存信度网的结构，就是要记录每一个结点的父结点。假设一个结点 X_i 的父结点数目为 k，在 n 个变量中，可以列出其所有可能的父结点集合，其数目为 $\binom{n}{k}$，即 n 个结点中任取 k 个结点的组合。在保存时，如果对这些父结点集合依次编上序号，就只需要保存一个结点的父结点的数目以及其父结点集合的序号。因此，保存信度网的结构所需的数据长度为（这里数据长度的单位是 Bit，所以以下的 log 表示 $\log_2$）：

$$DL_{Stru}(G)=\sum_{i=1}^{n}\left(\log k+\log\binom{n}{k}\right)$$

对于一个变量 X_i，用 $|Pa_i|$ 表示其父结点集合的取值组合的数目，用 $|X_i|$ 表示变量的取值数目，则其条件概率表所包含的项数为 $|Pa_i||X_i|$。根据概率的归一性原理，只需要保存其中的 $|Pa_i||X_i-1|$ 项就行了。对于条件概率表中的每一项，Friedman 和 Yakhini [1996]指出，需要 $1/2\cdot\log N$ 个 Bit 位来表示。其中 N 为实例数据集合中实例的个数。则保存条件概率表所需的数据长度为：

$$DL_{Para}(G)=\sum_{i=1}^{n}1/2\cdot|Pa_i|(|X_i|-1)\log N$$

2）实例数据的压缩长度　这里采用 Hafferman 编码。对一个实例，根据它的分布（即出现的频率）确定编码长度。其概率越大，则编码长度越小；概率越小，则其编码长度越大。Cover 和 Thomas[1991]指出在这种编码中，一个实例 C_i 的二进制编码长度近似等于 $\log 1/P(C_i)$。因此实例数据编码压缩后的长度为：

$$DL_{data}(C|G)=-\sum_{i=1}^{N}\log P(C_i|G)$$

由于各个实例相互独立，且实例数据集合是完整的，所以：

$$\begin{aligned}DL_{data}(C|G)&=-\sum_{i=1}^{N}\log P(C_i|G)\\&=-\log P(C|G)\\&=-\log\prod_{i=1}^{n}\prod_{Pa_i}\prod_{X_i}P(X_i|Pa_i)^{N_{ijk}}\\&=-\sum_{i=1}^{n}\sum_{X_i,Pa_i}N*\hat{P}(X_i,Pa_i)\log P(X_i|Pa_i)\\&=N\sum_{i=1}^{n}H(X_i|Pa_i)\end{aligned}$$

其中，$H(X|Y)=-\sum_{x,y}P(x,y)\log P(x|y)$ 称为变量 X 对于变量 Y 的条件熵，对于完整的实例数据集合，条件熵可以通过对实例的个数统计来求出。

综上所述，一个结构的 MDL 测度为：

$$\begin{aligned}Score_{MDL}(G:C)&=DL_{Stru}(G)+DL_{Para}(G)+DL_{data}(C|G)\\&=\sum_{i=1}^{n}(\log n+\log(\binom{n}{|Pa_i|}))+\\&\frac{1}{2}\sum_{i=1}^{n}|Pa_i|(|X_i|-1)\log N+N\sum_{i=1}^{n}H(X_i|Pa_i)\end{aligned}$$

3. 两种测度的比较

BDe 测度和 MDL 测度是以不同的理论基础推导出来的，它们之间的关系可以从以下几个方面分析：1）当实例数据的数目 N 非常大时，两个测度的计算结果趋向相同。2）在计算一个结构的 BDe 测度时，必须先计算其参数的先验分布，而在 MDL 测度中则不需要。所以，计算 MDL 测度要比计算 BDe 测度简单。然而由此造成的问题是，由于 MDL 测度没有用到先验知识，其学习结果的正确性完全依赖于实例数据集合，这就要求实例数据的数目必须很大，并不能出现大的偏差。而使用 BDe 测度学习时，通过对参数的先验值的估计，其依赖性比 MDL 测度要低，有时甚至可以纠正实例数据中的偏差。3）BDe 测度中没有明确地包含结构复杂性指标，在某些情况下，该测度会倾向于选择较为复杂的信度网结构；而在 MDL 测度中，明确地将结构复杂性作为一个指标，因此倾向于选择较简单的信度网结构，计算的结果更简单，更容易被接受。4）对于两个具有等价性的结构（独立性等价），建立在似然等价性假设上的 BDe 测度值是相同的，而它们的 MDL 测度值却不能保证相同。因此，对于非因果型信度网的学习，两个测度都可以使用，而对于因果型信度网的学习，因为 BDe 测度无法区分两个独立性等价的信度网结构，所以无法直接用于因果型信度网的学习。

三、测度的可分解性及启发式搜索算法

信度网的结构学习问题是一个 NP 问题，所以在实际的计算中，并不是对所有的结构分别计算其测度值，再进行比较取最优，而是采用启发式搜索算法，以选定的测度为指导，在可能的拓扑结构空间中进行搜索来获取信度网的拓扑结构。常用的启发式搜索算法有“瞎子爬山法”和模拟“退火”法等。应用启发式搜索算法的关键是测度函数必须具有可分解性。

定义2[3]　一个测度函数是可分解的，是指结构的测度值可以依据结构中所包含的依赖关系，分解为多个项相加的形式。其中的每一项是关于一个节点与其所有父节点所构成的简单图形的测度，并将每一项称

为本地测度函数，即：

$$Score(G)=\sum_{i=1}^{n}Score(X_i,Pa_i)$$

利用测度的可分解性，可以将计算限制在图形结构的局部，极大地简化了搜索算法的计算量。根据测度可分解性的定义，通过观察可知，BDe 测度和 MDL 测度都具有可分解性。

1. 最陡爬山算法

最陡爬山法的执行过程很象一个瞎子爬山的过程，由于看不见山峰在哪里，他只能用手杖测量四周的地形，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。最陡爬山算法如下：

第一步 初始化。确定一个初始结构，作为信度网的拓扑结构，并求出该结构的测度值。

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；改变一有向边的方向等），要求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。对 E 中的每一个修改 e，求出修改后该结构测度值的变化量 $\Delta(e)$。

第三步 如果所有的 $\Delta(e)$ 都小于0，则停止，此时的拓扑结构就是所求；如果不是所有的 $\Delta(e)$ 都小于0，则取 $\Delta(e)$ 最大的修改，修改当前的结构，并计算出该结构的测度值，转到第二步，继续。

在上面的算法中，第二步求解 $\Delta(e)$ 时，利用了测度函数的可分解性，减少了计算量。例如，假设修改 e 只是给节点 X_i 添加了一条指向它的有向边，此时只有节点 X_i 的本地测度发生了变化，其它的节点的本地测度都没有变化，所以只要求出节点 X_i 本地测度，就可以计算出 $\Delta(e)$，并不需要对所有的结点重新进行计算。

爬山算法存在的问题是，并不能保证求得的结果是全局最优的，很可能只是局部最优。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，可以任意地改变当前的结构，再以该结构作为初始结构，利用最陡爬山法继续学习，反复进行数次，最后可能获得全局最优的结构。

2. 模拟退火算法

"模拟退火法"是 Metropolis[1953]等提出的，该方法可以有效地解决局部最优问题。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，能够较自由地移动，从而跳过那些比较浅的局部最小。模拟退火算法如下：

第一步 初始化。首先确定一个信度网的初始拓扑结构（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，且置循环变量 $t=0$。

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；改变一有向边的方向等），要求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，修改后结构测度值的变化量为 $\Delta(e)$。求出以下的表达式的值：$P=\exp(\Delta(e)/T_0)$

第三步 如果 $p>1$，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 α 次。如果在 α 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 $t=t+1$。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 $P=exp(\Delta(e)/T_0)$ 可以看出，当随机选取的修改 e 引起的测度值的变化量 $\Delta(e)$ 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测度值增大的方向前进；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测度值减小的方向前进，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大点，从而解决了局部最大问题。

四、进一步的研究

在实际的信度网结构学习中，遇到的实例数据往往是不完整的。从不完整的实例数据学习信度网的结构要复杂得多，这一问题也是当前信度网研究的一个热点。

利用前面提到的 BDe 测度和 MDL 测度，配合近似的参数估计方法（如 EM 算法、梯度上升算法等），仍然可以学习信度网的结构。其基本过程为：

首先，对每一个可能的结构 G_i：

1）利用一些参数的近似计算方法（如 EM 算法、梯度上升算法及 Gibbs 抽样仿真算法等）计算出对应的参数 $\hat{\theta}_i$；

2）计算不完整实例数据的充分统计量的期望值，从而获得完整的实例数据，利用该完整的实例数据计

算结构 G_i 的测度 $Score(G_i:C)$。

最后，从中选取测度值最优的结构 G_i 作为信度网的结构。

可以看出，以上过程的计算量很大，一般当可能结构的数目很小时才使用。它存在如下的问题：①由于对每个可能结构都要利用参数的近似算法，所以算法的复杂性很高。②测度不再具有可分解性，所以无法使用启发式搜索算法，计算量很大。

Friedman[6,7]提出了一种不完整实例数据的结构学习算法—SEM 算法（Structure Expection Maimumal）。其基本思想是：不直接计算结构的测度，而是确定一种新的测度作为优化指标，该新测度满足两个条件：①新测度最优的结构，原先的测度也最优；②新测度具有可分解性质。期望测度就是符合上述要求的新测度：设当前结构为 G_i，利用近似的参数估计方法（如 EM 算法、梯度上升算法等）可以确定该信度网的参数 $\vec{\theta}_i$，从而获得信度网 $B=\{G_i,\vec{\theta}_i\}$。在该信度网上，利用信度网的推理算法，就可以估计出不完整实例数据中所有没有观测值的变量的值，从而获得一个完整的实例数据库。对于完整的实例数据库，可以容易地计算出任一个结构 G_i 的测度（如 BDe 测度或 MDL 测度），该测度称为 G_i 相对于 G_i 的期望测度，表示为 $E[Score(G_i:C)|G_i,C]$。可以证明，一个结构的期望测度越优，其测度也越优；同时可以看出，由于利用了完整的实例数据，所以期望测度具有可分解性，SEM 算法如下：

第一步（初始化）：确定一个信度网的初始拓扑结构 G_i（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）。

第二步：利用近似计算算法（如 EM 算法、梯度上升算法、Gibbs 仿真算法等），计算结构 G_i 对应的参数 $\vec{\theta}_i$。

第三步（Expectuon Step）：利用启发式搜索算法（如"瞎子"爬山法、模拟退火法等），计算所有可能结构的期望测度 $E[Score(G_i:C)|G_i,C]$。

第四步（Maximimal Step）：选择期望测度最大的结构 G_i。

第五步：如果 $E[Score(G_i:C)|G_i,C]\approx E[Score(G_i:C)|G_i,C]$，则停止算法，结构 G_i 即为求得的结构；否则，将 G_i 作为 G_i，转第二步继续执行。

SEM 算法类似于实例数据不完整时计算条件概率表的 EM 算法，它避免了对每个可能结构都估计其参数，在每一次循环中，只对期望测度最大的结构进行参数估计，极大地减少了所需的计算量。另外，在算法的第三步，利用期望测度的可分解性，采用了启发式搜索算法来计算各个可能结构的期望测度，进一步减少了计算量。所以 SEM 算法可以高效地利用不完整实例数据库来学习信度网结构。

（下转第65页）

（上接第103页）

已证明 Petri 网的模拟能力与图灵机等价，Petri 网应用已涉及计算机学科的各个领域，如：线路设计、网络协议、软件工程、人工智能、形式语义、操作系统、并行编译、数据库管理等等。

Petri 网是一种可用图形表示的组合模型，具有直观、易懂和易用的优点，有严格定义的数学对象，既可用于静态结构的分析，又可用于动态结构的分析，有人预测它的发展必将为信息论奠定坚实的理论基础。

Petri 网的目的是为了克服有限自动机用系统的全局状态刻画变化所带来的局限性：首先，尽管有限状态机的状态元素只有有限个，当这个数目足够大时，全局状态仍然不是实时可知的，从而无从实时确定什么变化可以发生；其次，就每个变化而言，它发生前后的两个状态并不能完整刻画变化（如：催化剂是某些化学反应必不可少的，却不能在反应前后的差异中体现出来）。

一般系统模型均包含两类元素：表示状态的元素（变量、角色、结构）和表示变化的元素（语句、活动、构造），表示变化的元素也叫变迁，变迁之所以能发生及发生影响完全取决于自然规律，不带主观性。

Petri 网没有任何象冯·诺依曼那种固定的控制。用 Petri 网描述的系统有一个共同的特征：系统的动态表现为资源的流动。Petri 网为 MAS 及其相关系统提供了设计分析工具。

结束语 MAS 和 Petri 网是研究"分布"、"并发"系统的理论基础，网络是实现这类系统的物理环境，其他相关概念最终将作为 MAS 相应能力的工具或支撑环境（见图1）。

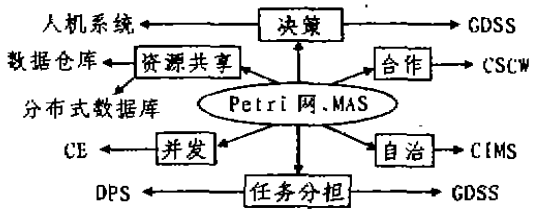

图1 MAS 及其相关概念之间的关系

参考文献

1 吴启迪. MAS 和 CSCW 在 CIMS 中. CIMS，1999，5(1)：1～6

2 沈锦涛. 群体决策支持系统与人工智能. 计算机科学，1995，22(1)：78～80

立基值，在网络系统活动时，根据利害关系设置权函数，修改变量值，比较与期望值的偏差检测入侵行为。

(3)基于专家系统的入侵检测技术　根据安全专家对入侵行为的分析经验形成一套推理规则，进而构成专家系统自动地检测入侵行为。入侵检测专家系统的适应性较强，其实现属于知识工作问题，随着入侵经验的积累，利用自学能力扩充和修正推理规则，提高入侵检测能力。

3. 防火墙技术

作为加强网络间访问控制的网络互连设备，防火墙是在内部网与外部网之间实施安全防范的系统，它保护内部网络免受非法用户的侵入，过滤不良信息，阻止信息资源的未授权访问。

防火墙是一种基于网络边界的被动安全技术，对内部未授权访问难以有效控制，因此较适合于内部网络相对独立，且与外部网络的互连途径有限、网络服务种类相对集中的网络。防火墙的实现技术主要有：数据包过滤，应用网关和代理服务等。

(1)包过滤技术　依据系统内事先设定的过滤逻辑，检查数据流中每个数据包，根据数据包的源地址、目的地址、所用的 TCP 端口与 TCP 链路状态等实施有选择的通过。包过滤技术的实现方式有：①路由设备除完成路由选择的数据转发外，还进行包过滤，这是较常用的方式；②在工作站上使用软件进行包过滤，价格较贵；③在屏蔽路由器上启动包过滤功能。

(2)应用网关技术　基于应用层协议，利用特别的网络应用服务协议分析过滤数据包，并形成相关的报告。应用网关一般运行在专用工作站上，对易登录、控制的网络系统实施严格控制，确保网络安全。

(3)代理服务(Proxy Server)技术　防火墙网关上建立的专用代码由服务器端程序和客户端程序组成，客户端程序与代理服务连接，代理服务与将访问的外部服务器连接。与包过滤技术和应用网关技术不同，代理服务技术的内部网与外部网间不存在直接连接，实现了防火墙内外计算机系统的隔离，同时，代理服务技术可实施较强的数据流监控、过滤、日志(Log)和审计(Audit)等服务。

利用防火墙技术可以解决网络层的安全问题，但是，防火墙防外不防内，不能识别用户的身份，进行身份认证、授权管理及控制数据的存取。

结束语　互联网是庞大的信息共享系统，其网络安全问题是一个综合性课题，涉及技术、管理、使用、立法等许多方面，包括网络系统的安全问题，以及信息数据的安全问题。

在信息时代，网络安全问题越来越重要。可以预言，网络安全将是21世纪世界十大热门课题之一。

参考文献

1 Stallings W. Network and Internetwork Security Principles and Practice. IEEE Computer Society Press, 1995

2 RFC 791, Internet Protocol, 1981

3 Hare C, Siyan K. Internet Firewalls and Network Security, 1996

4 Anonymous. Maximum Security, Sams. net Publishing, 1997

5 Farley M, Stearns T. LAN Times Guide to Security and Data Integrity. McGraw-Hill Inc., 1996

6 Escamilla T. Intrusion Detection. John Wiley & Sons Inc., 1998

(上接第87页)

主要参考文献

1 Verma T, Pearl J. Equivalence and synthesis of causal models. In: Proc. of Sixth Conference on Uncertainty in AI, Boston, MA, Morgan Kaufmann, 1990. 220～227

2 Heckerman D. Learning Bayesian Network: The Conbination of Knowledge and Statistical Data. [Technical Report MSR-94-09]. 1994

3 Heckerman D. A tutorial on learning with Bayesian networks. [Technical Report, MSR-TR-95-06]. 1995

4 Heckerman D, et al. Real world applications of Bayesian networks. Communications of the ACM, 1995, 38

5 Rissanen J. Stochastic Complexity in Statistical Inquiry. World Scientific, River Edge, NJ, 1989

6 Frieman N. Learning Bayesian networks in the presence of missing values and hidden variables. In: ML'97, 1997

7 Frieman N, The Bayesian structure EM algorithm. In: Fourteenth Conf. on Uncertainty in AI, 1998