

一种基于结构分解的影响图模型选择算法^{*})

姚宏亮 王浩 张佑生 方宝富

(合肥工业大学计算机与信息学院 合肥 230009)

摘要 针对概率网络模型选择的数据依赖性和计算复杂性,以及影响图模型的非概率关系学习问题,通过对影响图结构进行分解,提出一种 PS-EM 算法实现影响图概率结构部分的模型选择,给出一种利用 BP 神经网络学习每个效用结点的局部效用函数来实现效用结构部分的模型选择方法。PS-EM 算法对 N. Friedman 的 SEM 算法进行改进,提出一种引入融合先验知识的 MDL 评分标准来降低传统 MDL 评分对数据的依赖性;通过将参数学习和结构评分分开计算提高计算效率。最后,在石油投机商模型上的结果显示,PS-EM 比 SEM 的时间性能要好,对数据依赖性要小,且效用部分的结构选择易于实现。

关键词 影响图, SEM 算法, BP 神经网络, MDL 评分

A Model Selection Algorithm of Influence Diagrams Based on Structural Decomposition

YAO Hong-Liang WANG Hao ZHANG You-Sheng FANG Bao-Fu

(Department of Computer Science and Technology, Hefei University of Technology, Hefei 230009)

Abstract The data dependency, computation complexity and non-probability relation problems are faced by model selection of Influence diagrams. Based on the decomposition of Influence diagrams, a PS-EM algorithm is presented for learning the probability structures of IDs, and a BP Neural Network is introduced to use learning local utility functions of the utility part. The SEM algorithm is improved by PS-EM algorithm. PS-EM algorithm presents a new MDL scoring which includes the prior knowledge of network structures for reducing the dependency on data, and learning parameters and scoring structure are separated for improving the computation efficiency. On the oil wildcatter model, the experiment results show that PS-EM algorithm is better the time performance and less the data dependency than criterion SEM algorithm, and the model selection of the utility part is easy to achieve.

Keywords Influence diagrams, Structural EM algorithm, BP neural network, Minimum description length scoring

由 Howard 在 1984 年提出的影响图已成为表示决策问题的一种重要模型^[1]。由于影响图的结构、参数与决策规则相联系^[2],特别是影响图和效用结点相联系的结构部分不是一种概率依赖关系,而是一种函数关系^[3],使得影响图的学习和贝叶斯网络的学习不同。近些年来,研究者集中于利用数据来学习概率分布,但对于效用函数的学习很少受到研究者关注^[4]。Heckerman 在设定已知效用结点结构的情况下,将决策结点看成随机结点,利用概率网络的学习方法来进行影响图的学习^[5];但对比较复杂的决策问题,效用结点的结构常难以获得。后来, U. Chajewska 给出一种对数据进行密度估计学习效用函数的方法^[6],但计算量太大。

根据影响图模型的结构特点,可以将影响图分解成两个部分:概率网络结构部分;影响图中与效用结点相联系的效用结构部分。进而,对这两个部分分别进行模型选择。

MDL 评分是概率网络结构学习的一种重要评分标准, MDL 评分不能表示网络结构的先验知识,致使其对数据的依赖性较大^[7],要求数据的规模较大,数据是完整的且没有太大的偏差,但 MDL 评分偏向于选取简单的网络结构,计算复杂度较低。文中提出一种融合先验知识的 MDL 评分,首次将网络的先验知识应用于 MDL 评分中,实现在增加很小时间复杂性的情况下,可明显降低对数据的信赖性。

概率网络结构学习算法面临随着网络结点的增加网络结构以指数增长的 NP-Hard 问题,因而网络结构学习大都使用启发式学习算法,如: EM 算法^[8],遗传算法^[9]等。Structural

EM(SEM)^[8]的评分函数具有可分解性,具有较好的计算性能;但 SEM 算法对于数据的依赖性较大,且 SEM 算法对于高维网络结构空间的搜索计算量很大。

由于神经网络可以通过训练学习数据中的模式来构造模型,适合利用非线性数据和含噪声数据来建模^[10],因而针对效用结构部分的非概率关系可利用 BP 神经网络来学习效用结构部分。但联合效用函数的学习在计算上是一个指数问题,文中将影响图的联合效用函数转换成局部效用函数的形式,然后对每个效用结点的局部效用函数利用 BP 网络进行学习,以提高效用结构部分的学习效率。

1 影响图模型 (Influence diagrams, IDs) 和影响图结构分解

1.1 影响图的定义

影响图是一个三元组 $ID=(G, P, U)$, 这里 $G=(V, E)$, 其中决策结点表示行为选择;随机结点表示和决策问题相关的随机变量;效用结点表示与行为、状态相联系的效用。设 $X=\{X_1, \dots, X_n\}$, $D=\{D_1, \dots, D_k\}$ 分别表示随机变量集和决策变量集, U 是一个局部效用函数集合 $U=\{U_1, \dots, U_k\}$ 。

结构策略模型 为每个决策结点 D_i 确定一个局部的决策规则 δ_i , δ_i 是 D_i 的父结点集 $Pa(D_i)$ 和行为选择之间的一个函数;影响图的策略 $\delta=\{\delta_i\}$, $i \in [1: k]$ 可定义成: $\Omega_{Pa(D)} \rightarrow \Omega_D$, 其中 $\Omega_{Pa(D)}$ 为决策结点的父结点集可能取值空间, Ω_D 表示决策结点的可能行为空间。

^{*}) 本文受国家自然科学基金(编号 60575023);合肥工业大学校科学研究基金(编号 050506F)资助。姚宏亮 博士生,研究方向:机器学习。

概率模型 在给定每个变量 X_i 的父结点集 $Pa(X_i)$ 时, 决策结点看成随机结点, 对于给定的决策规则 δ_i , 决策结点 D_i 有如下条件概率分布:

$$P_{\delta_i}(D_i | Pa(D_i)) = \begin{cases} 1 & \text{当 } \delta_i(Pa(D_i)) = D_i \\ 0 & \text{否则} \end{cases}$$

随机变量和决策变量的联合概率分布可表示为:

$$P(X, D) = \prod_{i=1}^n P(X_i | Pa(X_i)) \prod_{i=1}^n P_{\delta_i}(D_i | Pa(D_i)) \quad (1)$$

效用模型 是一个联系决策结点和状态变量的函数。每个效用结点 U_i , 有一个和 U_i 父结点集 $Pa(U_i)$ 相联系的局部效用函数 $U_i(Pa(U_i))$ 。各个局部效用和为:

$$U(X, D) = \sum_{i=1}^n U_i(Pa(U_i)) \quad (2)$$

联合(1)和(2)式, 可计算全局决策 U_δ :

$$U_\delta = \sum_{X, D} P(X, D) \sum_{i=1}^n U_i(Pa(U_i)) \quad (3)$$

1.2 影响图结构的分解

设 V_B 是从影响图结点集 V 中去掉所有效用结点所得到的结点集, 因而 V_B 可以看成是一个在给定决策规则 $\delta = (\delta_1, \dots, \delta_k)$ 下的概率网络结点集。

效用结点和随机结点之间的关系不是概率依赖关系, 而是一种函数关系, 所以不能用学习概率结构关系的方法来学习效用结构; 另外, 容易从影响图结构中分离出效用结点, 这是因为效用结点没有孩子结点, 使得这种分解不会破坏网络的内部结构, 并可使网络的结构变得简单而易于实现模型选择。

2 融合结构先验知识的 MDL 评分

MDL 评分是选取对实例数据编码的描述长度和对网络结构编码的描述长度之和最小的网络结构。将网络结构的先验知识融合到 MDL 评分中, 可降低模型对数据的依赖性。

融合结构先验知识的网络结构总描述长度包含四部分: 样本数据的描述长度; 现在生成的网络结构描述长度; 已存在的网络结构描述长度; 复杂结构的惩罚项。设 $X = \{X_1, \dots, X_n\}$ 为随机变量集合, $S = \{x^1, \dots, x^N\}$ 是 X 的样本集, 对于每个 $X_i \in X$ 有一个可能的取值集 $\{x_i^1, x_i^2, \dots, x_i^{n_i}\}$, 其中 n_i 为可能取到的值的个数。

(1) 实例数据描述长度。设样本集 S 是完备的, $N_X(X=x)$ 表示 S 中 $X=x$ 样本的个数, 数据描述长度为:

$$H(X_i | Pa(X_i)) = - \sum_{X_i \in XX_i} \sum_{Pa(X_i)} N(X_i, Pa(X_i)) \log(\theta_{X_i | Pa(X_i)}) \quad (4)$$

其中 $H(X_i | Pa(X_i))$ 是信息熵。

当数据不完备时, $\theta_{X_i | Pa(X_i)}$ 表示成期望形式:

$$\theta_{X_i | Pa(X_i)} = \frac{E[N(X_i, Pa(X_i)) | X_i = x_i^j]}{E[N(Pa(X_i)) | X_i = x_i^j]}$$

(2) 设变量的个数为 n , 存储一个数字值所需要的二进制位数是常量为 d , 其公式为:

$$\sum_{X_i \in X} [k_i \log_2(n) + d(n_i - 1) \prod_{X_j \in Pa(X_i)} n_j] \quad (5)$$

其中, n_i 是 X_i 可能的状态数, k_i 是 X_i 父结点的个数。

(3) 已存在网络结构和现在生成结构差异的描述长度。设有一个现在生成的网络结构 B_N 的描述, 已存在的网络结构 B_E 仅描述其和现在生成的网络结构的不同方面。这些不同在于三个方面: 逆向边 (B_E 和 B_N 中都有但方向不同); B_N 中有, 在 B_E 中没有的边; B_E 中添加的边。对于 B_E 中的每一个结点 X_i , 在给定 B_N 时, 设 r_i, a_i, m_i 为分配给它的逆向, 添

加和丢失的边的个数。已存在网络结构和现在生成结构差异的描述长度可表示为:

$$\sum_{X_i \in X} [(r_i + m_i + a_i) \log_2[n(n-1)]] \quad (6)$$

其中: r_i, a_i, m_i 中的每个只要对 X_i 和 $Pa(X_i)$ 进行检测就可计算出; 网络中有 n 个结点, 则可能的边数为 $n(n-1)$ 。

已存在的网络结构是从以前的学习中所获得的, 是网络结构的先验知识; 用于学习的样本数据和存在的网络结构是相互独立的, 这使得数据的规模和数据的丢失率对于网络结构学习所产生影响将减小。已存在网络结构和现在生成网络结构差异的描述长度可看成是对使用高丢失率的样本学习网络结构而引起已存在网络结构显著变化的一种惩罚。

(4) 复杂结构的惩罚项

$$\sum_{X_i \in X} \frac{\log N}{2} B(X_i, Pa(X_i) | X_i = x_i^j) \quad (7)$$

其中 $B(X_i, Pa(X_i) | x_i^j)$ 表示描述 $P(X_i | Pa(X_i))$ 所需参数个数。复杂结构的惩罚项使 MDL 评分趋向于寻找一个结构较简单的网络, 实现网络的精度和复杂度的均衡。

网络结构 B 的总描述长度是上述四个描述长度之和:

$$Scor(B; S) = \sum_{X_i \in X} Scor_{X_i}(\theta_i, Pa(X_i); S) \quad (8)$$

3 影响图模型选择算法

3.1 概率网络部分的 PS-EM 模型选择算法

利用 EM 来确定模型的最大参数, 并在参数学习过程中对网络结构进行选择。

3.1.1 PS-EM 模型选择算法的评价标准

(1) 概率网络部分 EM 参数评分

设 $X = \{X_1, \dots, X_n\}$ 为网络中结点变量集合, 假设样本集 $S = \{x^1, \dots, x^N\}$ 有 N 条记录, 对于每个变量 $X_i \in X$ 有一个可能的取值集 $x_i = \{x_i^1, x_i^2, \dots, x_i^{n_i}\}$, 其中 n_i 为可能取到的值的个数。设有一系列的模型 $G = \{M_0, \dots, M_n\}$, 每个模型 $M_i \in G$ 是由向量 θ_M 参数化, 每个 θ_M 值定义了一个 X 上的概率分布 $P(\cdot; M, \theta_M)$, 下面用 θ 代替 θ_M 。

对于给定模型 M^* 和 X 的值, 确定模型的参数, 是要寻找一个参数向量 $(\theta; M^*, S)$ 使下面的评分函数最大化:

$$Q(\theta; M^*, S) = \sum_{X_i \in X} Q_{X_i}(\theta; M^*, X_i = x_i^j) = \sum_{X_i \in X} E[N(X_i, Pa(X_i)) | X_i = x_i^j] \log_2(\theta_{X_i | Pa(X_i)}) \quad (9)$$

(2) 概率网络部分的结构评分

对给定参数 θ^* 和样本集 S 的模型选择, 是寻找一个 $(M; \theta^*, S)$ 使融合先验知识的 MDL 评分取最小值的模型:

$$Scor(M; \theta^*, S) = - \sum_{X_i \in XX_i} \sum_{Pa(X_i)} N(X_i, Pa(X_i)) \log(\theta_{X_i | Pa(X_i)}^*) + \sum_{X_i \in X} [(r_i + m_i + a_i) \log[n(n-1)]] + \sum_{X_i \in X} [k_i \log(n) + (n_i - 1) \prod_{X_j \in Pa(X_i)} n_j] + \sum_{X_i \in X} \frac{\log N}{2} B(X_i, Pa(X_i) | X_i = x_i^j) \quad (10)$$

3.1.2 PS-EM 模型选择算法的描述

首先利用 EM 算法来实现参数最大化, 然后利用融合结构先验知识的 MDL 评分来进行模型选择。具体算法描述如下:

给定初值模型结构和参数选择 M^0 和 θ^0
 loop $n=0, 1, \dots$ 直至算法收敛
 { loop $l=0, 1, \dots$ 直至算法收敛或者 $l=l_{max}$
 { 使得 $\theta^{n,l+1} = \arg \max_{\theta} Q(\theta; M^n, \theta^{n,l}, S)$;
 $\theta^{n+1} = \theta^{n,l+1}$;
 将 $Scor(M; \theta^n, S)$ 和 $Scor(M; \theta^{n+1}, S)$ 和中得分最小的模型赋给

M^{n+1} ;
将 M^{n+1} 模型的参数赋给 $\theta^{n+1,0}$

3.2 效用结构部分的 BP 神经网络学习

影响图联合效用结构的学习是一个指数问题,通过学习各个局部效用函数有效实现联合效用结构的学习。

(1) 局部效用函数的表示

每个效用结点和一个局部效用函数相联系,一个效用结点 U_j 的效用函数可表示如下:

$$U_j(Pa(U_j)) = f_j(X_1, \dots, X_N) = w_j^1 X_1 + \dots + w_j^N X_N \quad (11)$$

其中初始设定 $Pa(U_j) = \{X_1, \dots, X_N\}$ 是效用结点 U_j 父结点集合; 权重 w_j^k 对应一个变量 $X_k \in Pa(U_j)$, 表示 X_k 对效用结点 U_j 影响的程度。 x_k 是 X_k 所取的值, D_j 的行为集为 $d_j = \{a_j^1, \dots, a_j^M\}$ 。

设变量 X_k 的值集合为 $x_k = \{x_k^1, x_k^2, \dots, x_k^i\}$, $x_{k,i}$ 表示给定决策结点 $D_j = a_j^i$ 和 X_k 的父结点集 $Pa(X_k)$ 的观察值时变量 X_k 的期望值, 即:

$$x_{k,i} = \sum_{j=1}^{n_k} x_k^j \times P(X_k = x_k^j | D_j = a_j^i, Pa(X_k)) \quad (12)$$

则在 $D_j = a_j^i$ 时, U_j 的效用函数可表示成:

$$U_j(Pa(U_j)) = \sum_{k=1}^N w_j^k x_{k,i} \quad (13)$$

(2) 学习局部效用函数的 BP 神经网络

给定 D_j 的决策序列 $\{d_j(1), \dots, d_j(T)\}$, 对于任意一个 $d_j(t)$ 有 $d_j(t) \in d_j, t=1, \dots, T$ 。 设初始局部效用函数中所有权重值为 0。 根据 $D_j = a_j^i$ 时 U_j 的效用函数为式(13), 则学习局部效用函数的 BP 网络模型可如图 1 所示。

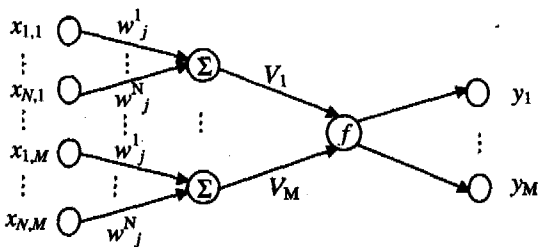


图 1 学习效用函数的 BP 神经网络

表示 BP 网络进行第 t 次决策的过程, 其中 $f(t)$ 是一个基于竞争的激励函数, 它由输出决策规则组成:

$$y_i(t) = \begin{cases} 1 & \text{对所有 } j \neq i, \text{ 有 } v_i(t) > v_j(t) \\ 0 & \text{否则} \end{cases} \quad (14)$$

其中, $i=1, \dots, m$; $v_i(t)$ 表示在 $D_j = a_j^i$, 偏差为 0 时有:

$$v_i(t) = \sum_{k=1}^N w_{k,j} x_{k,i}$$

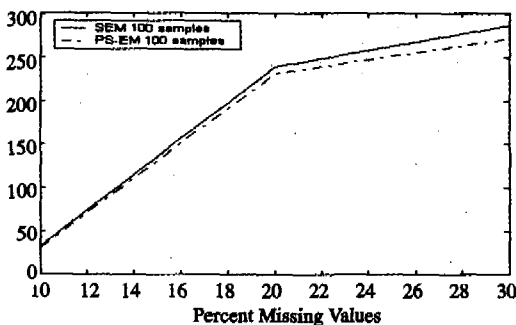


图 3 PS-EM 算法和 SEM 算法的时间性能

令 $z_i(t)$ 是 D_j 在第 t 次时的期望输出值, 为:

$$z_i(t) = \begin{cases} 1 & d_j(t) = a_j^i \\ 0 & \text{否则} \end{cases} \quad (15)$$

设 $e_i(t)$ 为决策误差 $e_i(t) = z_i(t) - y_i(t)$ 。

根据 Delta 规则对 U_j 的权重进行调整, 权重的调整与误差和输入单元的积成比例, 对每个输入单元进行标准化, 则在第 t 次决策时的权重调整规则为:

$$\Delta w_j^k(t) = \eta \sum_{i=1}^M e_i(t) \frac{x_{k,i}}{\sqrt{\sum_{j=1}^M \sum_{i=1}^N (x_{j,i})^2}} \quad (16)$$

其中, $k=1, \dots, N, \eta$ 是一个表示学习速率的常量。

4 算法比较和实验仿真

石油投机商模型是一个投机商决定是否钻井开采石油问题, 该模型由如下的结点组成: 随机结点, Oil 和 Result 为测试结果; 决策结点, Drill 和 Test; 效用结点, UD 和 UT。

对于该模型, 在给定初始的链式网络结构, 采用随机生成的同一组样本, 取迭代步骤为 30 的情况, 并以贝叶斯网络工具包(FullBNT)为基础, 以 Matlab 为运行环境。

图 2 表示 PS-EM 算法和 SEM 算法对贝叶斯网络部分进行模型选择时, 样本丢失率和学习所得模型的对数损失(Log-Loss)之间的关系。 学习所得模型的对数损失:

$\text{Log-Loss} = \sum_{i=1}^N \log(P_B(x^i))$ 表示学习所得模型 B 对样本集 S 拟合的精度, 对数损失值越高越好。 图 2 的实验结果表明对于相同的样本规模和相同的丢失率, PS-EM 算法比 SEM 算法更精确; 特别是随着数据丢失率的增大, 样本规模的减小, PS-EM 算法和 SEM 算法的精度差距会加大。 这说明 PS-EM 算法比 SEM 算法对数据的依赖性要小。

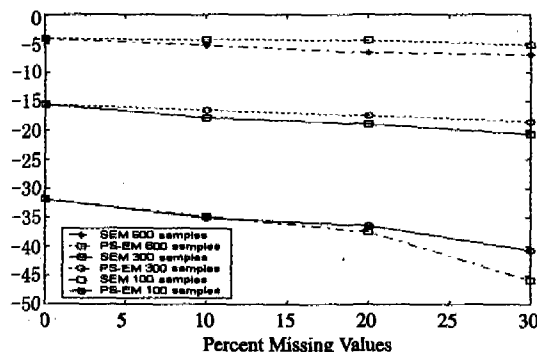
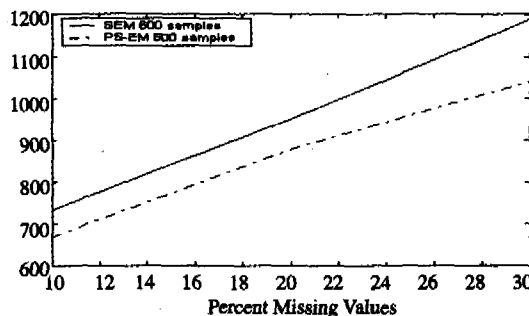


图 2 PS-EM 算法和 SEM 算法的对数损失



Springer-Verlag, 1996. 121~130

- 5 Clark P, Niblett T. The CN 2 induction algorithm. *Machine learning*, 1989(3), 261~283
- 6 Cover T M, Thomas J A. *Elements of Information Theory*. New York: John Wiley, 1991
- 7 Domingos P, Pazzani M. Beyond independence: Conditions for the optimality of the simple Bayesian classifier. In: *Proceedings of the Thirteenth International Conference on Machine Learning San Francisco, CA: Morgan Kaufmann, 1991. 105~112*
- 8 Duda R O, Hart P E, Stork D G. *Pattern classification and scene analysis*. New York: John Wiley, 2001
- 9 Frank E, Hall M, Pfahringer B. Locally Weighted Naive Bayes. In: *Proceedings of the Conference on Uncertainty in Artificial Intelligence, 2003. 249~256*
- 10 Friedman N, Goldszmidt M. Building classifiers using Bayesian networks. In: *Proceedings of the Thirteenth National Conference on Artificial Intelligence, Menlo Park, CA: The AAAI Press, 1996. 1277~1284*
- 11 Han J W, Kamber M. *Data Mining: Concepts and Techniques*. Academic Press, 2000
- 12 Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence, San Mateo, CA: Morgan Kaufmann, 1995. 1137~1143*
- 13 Kohavi R. Scaling up the accuracy of naive-Bayes classifiers; A

- decision-tree hybrid. In: *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*. Menlo Park, CA: The AAAI Press, 1996. 202~207
- 14 Kononenko I. Semi-naive Bayesian classifier. In: *Proceedings of European Conference on Artificial Intelligence, 1991. 206~219*
- 15 Langley P, Iba W F, Thompson K. An analysis of Bayesian classifiers. In: *Proceedings of the Tenth National Conference on Artificial Intelligence, Menlo Park, CA: The AAAI Press, 1992. 223~228*
- 16 Langley P, Sage S. Induction of selective Bayesian classifiers. In: *Proceedings of the Tenth Conference on Uncertainty in Artificial Intelligence, Seattle, WA: Morgan Kaufmann, 1994. 339~406*
- 17 Pazzani M J. Constructive induction of Cartesian product attributes. In: *Proceedings of the Conference, ISIS'96: Information, Statistics and Induction in Science, Singapore: World Scientific, 1996. 66~77*
- 18 Sahami M. Learning limited dependence Bayesian classifiers. In: *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, Menlo Park, CA: The AAAI Press, 1996. 334~338*
- 19 Xie Z P, Hsu W, Liu Z T, et al. SNNB: A Selective Neighborhood Based Naive Bayes for Lazy Learning. In: *Proceedings of the Sixth Pacific-Asia Conference on KDD, 2002. 104~114*
- 20 Zheng Z, Webb G I. Lazy learning of Bayesian rules. *Machine Learning*, 2000, 41(1), 53~84

(上接第 135 页)

图 3 表示取迭代步骤为 30, 样本规模为 100 和 600 时, 不同的丢失率情况下的 PS-EM 算法和 SEM 算法的时间性能比较。实验结果表明, 对于相同规模样本和相同的数据丢失率, PS-EM 算法比 SEM 算法所花费时间要少, 虽然融合结构先验知识的 MDL 评分中增加了先验结构描述长度的计算时间; 另外, 还说明数据的丢失变率会对两种算法的时间性能都产生显著的影响, 但从曲线的变化率可以看出数据丢失率对 PS-EM 算法的影响相对小一些。

在数据丢失率为 10% 时, 取样本数为 300 和 600 时, PS-EM 算法和 BP 网络方法联合学习石油投机商模型所得到的影响图结构如图 4 和图 5 所示。实验结果表明, 在数据丢失率为 10% 时, 取样本数为 600 时, 学习所得到的影响图结构已和标准结构相同。在数据是完备的情况下, 利用 300 个样本就可以得到标准结构。

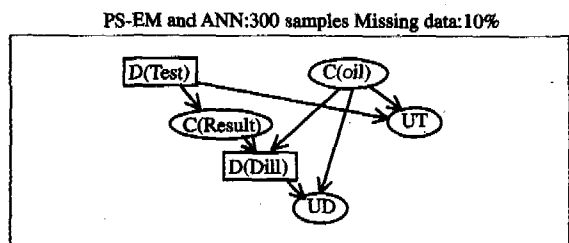


图 4 PS-EM 和 BP 联合学习石油投机商模型的结构

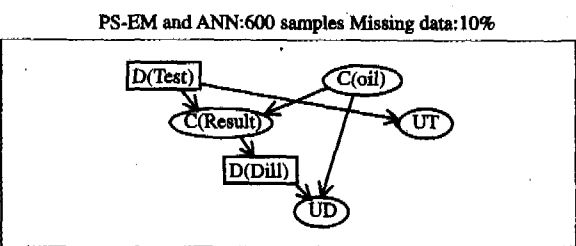


图 5 PS-EM 和 BP 联合学习石油投机商模型的结构

结束语 本文通过将影响图分解成概率网络结构部分和效用部分来分别进行结构学习。提出一种 PS-EM 算法用于概率网络部分的模型选择, PS-EM 算法是通过引入融合结构先验知识的 MDL 评分, 来降低 MDL 评分对数据的依赖性; 并通过将参数学习和结构评分分开计算提高模型选择的计算效率。效用结构部分学习是通过利用 BP 网络来学习每个效用结点的局部效用函数来实现。

参 考 文 献

- 1 Howard R A, Matheson J E. Influence diagrams [R]. *Readings on the Principles and Applications of Decision Analysis*, 1984. 719~792
- 2 Dawld A P. Influence Diagrams for Causal Modeling and Inference [J]. *International Statistical Review*, 2002, 70(2); 161~189
- 3 Pettersson. Application of the utility function method for behavioral organization in a locomotion task [J]. *Evolutionary Computation, IEEE Trans*, 2005, 9(5); 506~521
- 4 Nielsen T D, Jensen F V. Learning a decision maker's utility function from (possibly) inconsistent behavior [J]. *Artificial Intelligence*, 2004, 160; 53~78
- 5 Heckerman D, Geier D. Learning Bayesian networks: the combination of knowledge and statistical data [J]. *Machine Learning*, 1995, 20(3); 197~243
- 6 Chajewska U, Koller D. Utilities as random variables: density estimation and structure discovery [C]. In: *Proc. of the Sixteenth Conference on Uncertainty in Artificial Intelligence, Stanford, CA, 2000. 63~71*
- 7 王双成, 苑森森. 具有丢失数据的贝叶斯网络结构学习研究 [J]. *软件学报*, 2004, 15(7); 1042~1048
- 8 LI Xiao-lin, Yuan Sen-miao. Learning Bayesian Networks Structures based on extending evolutionary programming [C]. In: *Proc. conference on Machine Learning and Cybernetics*, 2004. 26~29
- 9 刘大有, 王飞, 卢奕南. 基于遗传算法的 Bayesian 网络结构学习研究 [J]. *计算机研究与发展*, 2001, 38(8); 916~922
- 10 Lampinen J. Bayesian approach for neural networks [J]. *Neural Networks*, 2001, 14; 256~274