

文章编号: 1001-0920(2007)05-0569-04

用于风险管理的贝叶斯网络学习

王双成^{1a}, 唐海燕^{1b}, 刘喜华²

(1. 上海立信会计学院 a. 信息科学系, b. 中国立信风险管理研究院,
上海 201600; 2. 青岛大学 经济学院, 山东 青岛 266071)

摘要: 结合专家知识和数据进行贝叶斯网络学习. 首先利用专家知识建立初始贝叶斯网络结构和参数; 然后基于变量之间基本依赖关系、基本结构和依赖分析方法, 对初始贝叶斯网络结构进行修正和调整, 得到新的贝叶斯网络结构; 最后将由专家和数据确定的参数合成为新的参数, 得到融合专家知识和数据的贝叶斯网络. 该方法可避免现有的贝叶斯网络学习过于依赖数据、对数据的数量和质量要求过高等问题.

关键词: 贝叶斯网络; 风险管理; 结构学习; 参数学习; 专家知识

中图分类号: TP181 **文献标识码:** A

Learning Bayesian networks in risk management

WANG Shuang-cheng^{1a}, TANG Hai-yan^{1b}, LIU Xi-hua²

(1a. Department of Information Science, 1b. China Lixin Risk Management Research Institute, Shanghai Lixin University of Commerce, Shanghai 201600, China; 2. Economic Institute, Qingdao University, Qingdao 266071, China. Correspondent: WANG Shuang-cheng, E-mail: wangsc@lixin.edu.cn)

Abstract: A new method of learning Bayesian networks is presented, which can effectively combine expert knowledge and data. Firstly, an initial Bayesian network structure is set up by using expert knowledge. Then, it is revised and regulated based on basic dependency relationship between variables, basic structure between nodes and dependency analysis method to obtain a new Bayesian network structure. Finally, two kinds of parameters got respectively by expert knowledge and data are fused to produce new parameters, and a Bayesian network combining expert knowledge and data is gained. This method can avoid the problems of depending on a large number of data with high quality in existing Bayesian network learning.

Key words: Bayesian network; Risk management; Structure learning; Parameter learning; Expert knowledge

1 引言

随着世界多极化和全球经济一体化进程的加快, 人类不仅受到技术发展不断变化的挑战, 而且面临着不断增长的系统性风险. 世界正进入一个不同于传统常态社会的风险社会, 风险已成为当今时代的主题之一.

贝叶斯网络^[1]是风险管理的有力工具之一^[2-5]. 其基础是贝叶斯网络学习, 包括结构学习和参数学习两部分, 结构学习是核心. 近十几年来, 基于数据的贝叶斯网络结构学习为主线, 相继发展了许多经典的算法. 这些算法都需要大量可靠的例子数据, 并进行复杂的运算. 可将这些算法大致分为两类: 打

分-搜索方法^[6,7]和依赖分析方法^[8]. 打分-搜索方法中的打分函数运算复杂程度和结构搜索空间大小都随变量的增加呈指数增长, 基于打分函数可分解性的局部搜索是 NP 难问题, 因此往往需要结点有序和启发式搜索, 得到的一般是局部最优结构. 依赖分析方法以双层邻域为条件集进行条件独立性检验, 导致大量的高维条件概率计算, 效率和可靠性很难得到保证. 在风险管理中, 积累大量可靠的例子比较困难, 因此这些方法在风险管理领域不具有实用性.

本文结合专家知识和数据进行贝叶斯网络结构和参数学习. 首先基于专家知识建立初始贝叶斯网络, 然后进行具有多项式复杂度的初始贝叶斯网络

收稿日期: 2006-04-24; 修回日期: 2006-06-14.

基金项目: 国家自然科学基金项目(60675036); 上海市重点学科基金项目(P1601); 上海市教委重点基金项目(05zz66).

作者简介: 王双成(1958—), 男, 上海人, 教授, 博士, 从事人工智能、风险管理等研究; 唐海燕(1962—), 男, 上海人, 教授, 博士生导师, 从事国际贸易、风险管理等研究.

结构修正和调整,并针对不同例子数量分别建立相应的近似结构调整方法;最后给出专家确定的参数和数据库统计参数的合成方法,可以避免现有贝叶斯网络学习方法存在的主要问题.

用 X_1, \dots, X_n 表示离散随机变量(简称变量),其值为 x_1, \dots, x_n ,数据库 D (或例子集)中具有 N 个记录(或例子).假设数据是独立地随机产生于概率分布 P ,概率模式中的变量和表示概率模式的图形模式中的结点有时不加区分.

2 初始贝叶斯网络学习

2.1 初始贝叶斯网络结构学习

对于选择的变量,初始贝叶斯网络结构(用 G 表示)学习主要依据专家的因果知识,其理论基础是因果马尔科夫假设^[9].即如果变量之间存在因果网络,那么这个因果网络就是贝叶斯网络.

在实际问题中,贝叶斯网络结构中一般包括3类变量(结点):第1类是受关注的假设变量(或目标变量);第2类是信息变量(能够观察到的变量);第3类是中间变量(或媒介变量).基于专家知识建立贝叶斯网络结构的过程如图1所示.

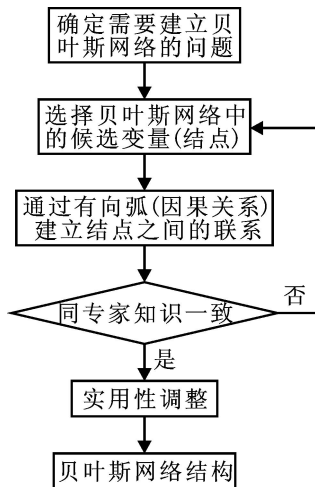


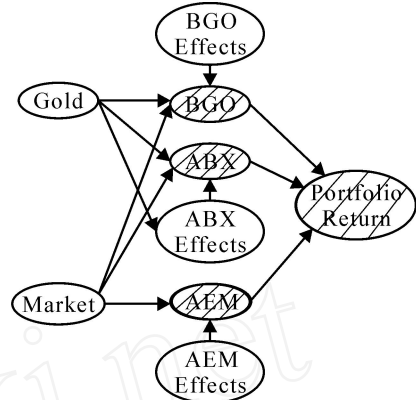
图1 基于专家知识建立贝叶斯网络结构的过程

在图1中,第2步~第4步往往需要多次反复,以确保贝叶斯网络结构与专家的知识结构一致.基于专家知识建立贝叶斯网络结构一般是不唯一的,它所刻画的是一个或一些专家的特有洞察问题视角、解决问题方式和知识结构.由专家确定的初始贝叶斯网络结构^[3]如图2(a)所示.

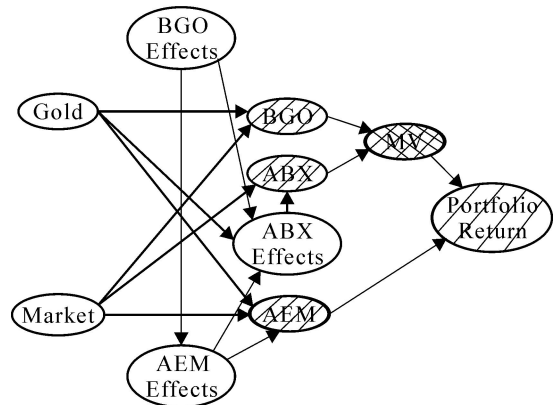
在图2(a)中,变量 Gold, Market, AEM Effects, BGO Effects, AEM Effects 是信息变量; BGO, ABX, AEM, Portfolio Return (分别表示黄金采矿业3支股票的利润和综合利润)是目标变量.

2.2 初始贝叶斯网络参数学习

根据专家知识确定参数是一项繁琐和困难的



(a) 专家确定的贝叶斯网络结构



(b) 修正与调整后的贝叶斯网络结构

图2 多证券投资的贝叶斯网络结构

工作,尤其是对具有大量参数的贝叶斯网络.由于专家一次确定大量的参数有一定困难,一般需要多次反复,不断地进行补充和调整.为避免偏差,有时还要附加一些约束条件.

设变量 $X_i \in \{X_1, \dots, X_n\}$ 的取值为 x_i^1, \dots, x_i^i , i 表示 X_i 的父结点集, $p_{ijk}^{(h)} = p(x_i^k | x_j^i, x_i^{(h)}, \dots, G)$ 是专家第 h 次补充和调整后的参数.其中: x_j^i 是 x_i 的第 j 个配置,表示专家关于离散变量参数的定量知识.

第 h 次修改前的参数为

$$p_{ijk}^{(h-)} = p(x_i^k | x_j^i, x_i^{(h-)}, \dots, G) \quad 0, \quad (1)$$

第 h 次修改后的参数为

$$p_{ijk}^{(h+)} = p(x_i^k | x_j^i, x_i^{(h+)}, \dots, G) \quad 0, \quad (2)$$

第 h 次的合成参数为

$$p_{ijk}^{(h)}(\dots, G) = \frac{p_{ijk}^{(h-)}(\dots, G) + p_{ijk}^{(h+)}(\dots, G)}{p_{ij}^{(h-)}(\dots, G) + p_{ij}^{(h+)}(\dots, G)} \quad 0. \quad (3)$$

3 贝叶斯网络的修正与调整

由专家知识建立的初始贝叶斯网络往往具有一定的主观性和局限性,使用客观数据进行修正与调整,可以实现互补.下面分别从结构和参数两方面进行修正与调整.

3.1 结构的修正与调整

在变量之间存在 3 种基本依赖关系^[10,11] (边) : 1) 传递依赖; 2) 非传递依赖; 3) 导出依赖. 结点之间对应地存在 3 种基本结构^[10,11]. 贝叶斯网络结构的修正与调整过程就是增加丢失的第 1 种边、去除多余的边和调整边方向的过程.

分别用 $I(X_i, X_j)$ 和 $I(X_i, X_j | X_{u_1}, \dots, X_{u_h})$ 表示变量 X_i 与 X_j 之间的互信息和以 X_{u_1}, \dots, X_{u_h} 为条件的条件互信息. 对于给定的小正数 (一般取 $\epsilon = 0.01$), 如果 $I(X_i, X_j | X_{u_1}, \dots, X_{u_h}) < \epsilon$ (或 $I(X_i, X_j) < \epsilon$), 则认为 X_i 与 X_j 之间条件(或边缘)独立. 用 L 表示根据专家知识所建立的贝叶斯网络结构边表, 其中的元素是四元组

$$(i, j, \delta, \sigma), j > i, i = 1, \dots, n, \delta = 0, 1, \sigma = 1, 2. \quad (4)$$

$\delta = 1$ 和 $\delta = 0$ 分别表示存在和不存在边 $X_i - X_j$, $\sigma = 1, 2$ 表示弱和强因果关系, G 是 G 的复制.

(1) 增加第 1 种边

对于不存在边的结点对, 依次进行互信息检验, 如果 $I(X_i, X_j) > \epsilon$, 则在 G 中增加边 $X_i - X_j$, 并修改边表 L . 这一过程最多需要 $n(n - 1)/2$ 次互信息计算.

(2) 去除第 2 种边

在增加第 1 种边的同时, 也会增加一些第 2 种边(有些第 2 种依赖也很强). 为此, 通过下述方法去除第 2 种边:

在 G 中, 用 $S_{X_i}(X_i, X_j)$ 和 $S_{X_j}(X_i, X_j)$ 分别表示 X_i 和 X_j 邻域中 X_i 和 X_j 链路(两个结点之间的无向通路)上的结点集, $S(X_i, X_j)$ 表示 $S_{X_i}(X_i, X_j)$ 和 $S_{X_j}(X_i, X_j)$ 中具有较少结点的结点集. 对于存在边的结点对 X_i 和 X_j , 设 $S(X_i, X_j) = \{X_k^{(i,j)}, \dots, X_t^{(i,j)}\}$, 如果存在 $j_0 (1 \leq j_0 \leq t)$, 使 $I(X_i, X_j | X_{j_0}^{(i,j)}, D) < \epsilon$, 则在 G 中删除边 $X_i - X_j$, 并修改边表 L (这一过程称为一阶条件独立性检验); 否则, 选取 $X_j^{(i,j)} = \arg \min_{X_k^{(i,j)}, k=1, \dots, t} \{I(X_i, X_j | X_k^{(i,j)}, D)\}$, 如果存在 $j_0 (1 \leq j_0 \leq t, j_0 \neq j^*)$, 使 $I(X_i, X_j | X_{j_0}^{(i,j)}, X_{j^*}^{(i,j)}, D) < \epsilon$, 则在 G 中删除边 $X_i - X_j$, 并修改边表 L (这一过程称为二阶条件独立性检验). 重新确定 $S(X_i, X_j)$, 由于产生第 2 种依赖的信息流往往能被少数结点所阻塞(多数情况是一个或两个结点), 大部分第 2 种边已被清除, $S(X_i, X_j)$ 将有很少的冗余结点. 如果 $I(X_i, X_j | S(X_i, X_j), D) < \epsilon$, 则在 G 中删除边 $X_i - X_j$, 并修改边表 L (这一过程称为完全条件独立性检验).

上述 3 个过程最多需要 $2n^3$ 次条件互信息计

算, 可依据例子数据的数量决定进行哪一层次的条件独立性检验.

(3) 调整边的方向

调整边的方向主要包括两个阶段: 第 1 阶段使用高效率的碰撞识别为部分边定向; 第 2 阶段基于变量之间的相对信息为其他边定向.

1) 使用碰撞识别调整方向. 对于选择的结点对 X_i 和 X_j , 在边表 L 中查询. 设 X_i 与 X_j 之间不存在边, 在 G 中与 X_i 和 X_j 可能形成 V 结构^[12] 的结点为 X_{m_1}, \dots, X_{m_s} , 对每个可能的 V 结构进行碰撞识别(汇聚识别). 对于给定的阈值 $\epsilon > 0$, 如果 $\frac{I(X_i, X_j | X_{m_h})}{I(X_i, X_j)} > (1 + \epsilon)^{1/h}$, 则 X_i, X_j 和 X_{m_h} 形成 V 结构, 定向为 $X_i \rightarrow X_{m_h}$ 和 $X_j \rightarrow X_{m_h}$, 这样可为部分边定向. 当与专家定向发生冲突时, $\epsilon = 2$ 的方向保留, $\epsilon = 1$ 的方向不保留. 这一过程最多需要 $n(n - 1)(n - 2)/2$ 次条件互信息计算.

2) 使用相对信息调整方向. 用

$$R(X_{m_1}, \dots, X_{m_w}, X_j | X_i) = \frac{H(X_i | X_{m_1}, \dots, X_{m_w}, X_j)}{H(X_i | X_{m_1}, \dots, X_{m_w})} | X_i \quad (5)$$

表示 X_j 对 X_i 提供的相对信息, 其中 H 是信息熵, 分母 $H(X_i | \dots)$ 是对多值倾向性的均衡.

对于任意两个变量 X_i 和 X_j , 在排除其他变量影响的情况下, 如果 X_i 对 X_j 比 X_j 对 X_i 提供更多的相对信息, 那么方向应由 X_i 指向 X_j .

用 $B(X_i)$ 表示 X_i 的父结点、子结点、子结点的父结点、与 X_i 通过无向边直接连接的结点集(不包括 X_j). 在 V 结构可识别的假设下, 可以证明 $B(X_i) \cap B(X_j)$ 是 X_i 的马尔科夫毯^[1]. 当 $B(X_i) \cap B(X_j)$ 中变量给定时, 除 X_i 和 X_j 之间的依赖外, 可以排除其他变量的影响.

$$R(B(X_i) \cap B(X_j), X_i \rightarrow X_j) > R(B(X_i) \cap B(X_j), X_j \rightarrow X_i), \quad (6)$$

而且 $X_i \rightarrow X_j$ 不产生环路, 定向为 $X_i \rightarrow X_j$; 如果产生环路, 则标记为待定向边.

$$R(B(X_i) \cap B(X_j), X_i \rightarrow X_j) < R(B(X_i) \cap B(X_j), X_j \rightarrow X_i) \quad (7)$$

的情况作类似处理.

$$R(B(X_i) \cap B(X_j), X_i \rightarrow X_j) = R(B(X_i) \cap B(X_j), X_j \rightarrow X_i), \quad (8)$$

标记为待定向边.

当与专家定向发生冲突时, 采用碰撞识别中的处理策略. 如果产生环路, 则标记为待定向边. 这一过程最多需要 $n(n - 1)/2$ 次相对信息计算. 待定向边使用专家知识或基于 MDL 的标准定向, 多数情

况能够完全定向. 即使存在待定向边, 数量也很少, 方向性也不会很强, 对网络结构影响不大. 综上所述, 贝叶斯网络结构修正与调整算法的时间复杂性为 $O(n^3)$.

3.2 参数的修正与调整

对于修正和调整后的贝叶斯网络结构, 需要结合数据对专家确定的初始贝叶斯网络参数进行修正和调整, 以得到新的参数.

新结构下的原参数为

$$p_{ijk}(x_i, G) = p(x_i | x_j, x_k, G) = \frac{p(x_1, \dots, x_n | G)}{p(x_1, \dots, x_n / -x_i, x_i | G)}$$

$$p_{ijk}(x_i, G) = \prod_{i=1}^n p(x_i | x_j, x_k, G). \quad (9)$$

新结构下的新参数为

$$p_{ijk}(x_i, D, G) = \frac{p_{ijk}(x_i, G) + N_{ijk}(D, G)}{p_{ij}(x_i, G) + N_{ij}(D, G)}. \quad (10)$$

4 实 验

根据网站 <http://www.norsys.com> 提供的 ALARM 网结构和概率分布表, 生成具有 5 000 个例子的模拟数据集用于实验. 以具有 10 个多余边、10 个丢失边和 10 个反向边的有向无环图作为初始贝叶斯网络, 分别使用 1 000、2 000、5 000 个例子进行一阶、二阶和完全条件独立性检验, 确定边的方向(所有方向都重新检验). 建立贝叶斯网络结构, 情况如表 1 所示.

表 1 贝叶斯网络结构修正与调整情况

独立性检验	增加边	丢失边	反向边	例子数
初始贝网	10	10	10	0
一阶检验	6	1	2	1 000
二阶检验	2	1	1	2 000
完全检验	0	1	0	5 000

从表 1 可以看出, 为保证可靠地进行一阶、二阶和完全条件独立性检验, 按例子数量来选择调整策略. 一般情况下, 二阶条件独立性检验方法能较可靠地进行贝叶斯网络结构调整, 并且具有较高的效率, 是一种非常实用的方法. 该方法也适合于建立高效贝叶斯网络分类器和特征子集选择. 本算法中完全条件独立性检验所需的例子数量明显少于其他贝叶斯网络结构学习方法, 在效率和可靠性方面有其优势.

下面是 3 个有代表性学习算法的 ALARM 网结构学习结果. Cheng 等^[8] 给出的算法使用具有 10 000 个记录的数据集进行结构学习, 结果丢失 2 条边, 没有多余边, 4 条边无法定向; Lam 等给出的

算法使用 MDL 作为局部打分函数进行打分-搜索, 结果丢失 3 条边, 2 条边定向错误; Spirtes 等给出的算法使用 10 000 个例子的数据学习, 结果丢失 3 条边, 有 2 条多余边.

5 结 语

本文建立了结合专家知识和数据的贝叶斯网络学习方法, 并对具有不同例子数量的情况, 分别制定了相应的学习策略. 当数据集较小时, 可能需要多次反复, 在专家与数据之间进行权衡和互补. 该方法能避免单独使用专家知识或数据所带来的主要问题, 适合于风险管理领域中贝叶斯网络学习, 为贝叶斯网络分类器学习和特征子集选择提供了有效实用的方法.

参考文献(References)

- [1] Pearl J. Probabilistic reasoning in intelligent systems: Networks of plausible inference [M]. San Mateo: Morgan Kaufmann, 1988.
- [2] Gemela J. Financial analysis using Bayesian networks [J]. Applied Stochastic Models in Business and Industry, 2001, 17(1): 57-67.
- [3] Kwabena A P. Operational risk management- implementing a Bayesian network for foreign exchange and money market [D]. Kumasi: University of Gottingen, 2005.
- [4] 阮彤, 冯东雷, 李京. 基于贝叶斯网络的信息过滤模型研究[J]. 计算机研究与发展, 2002, 39(12): 34-41. (Ruan T, Feng D L, Li J. An information filtering model based on Bayesian networks[J]. J of Computer Research and Development, 2002, 39(12): 34-41.)
- [5] 冯力, 管晓宏, 郭三刚, 等. 采用规划识别理论预测系统调用序列中的入侵企图[J]. 计算机学报, 2004, 27(8): 76-84. (Feng L, Guan X H, Guo S G, et al. Plan recognition based method for predicting intrusion intentions of system call sequences [J]. Chinese J of Computers, 2004, 27(8): 76-84.)
- [6] Cooper G F, Herskovits E. A Bayesian method for the induction of probabilistic networks from data [J]. Machine Learning, 1992, 9(4): 309-347.
- [7] Heckerman D, Geiger D, Chickering D M. Learning Bayesian networks: The combination of knowledge and statistical data [J]. Machine Learning, 1995, 20(3): 197-243.
- [8] Cheng J, Bell D, Liu W R. Learning Bayesian networks from data: An efficient approach based on information theory[J]. Artificial Intelligence, 2002, 137(1/2): 43-90.

(下转第 576 页)

引理 4^[5] 最优鲁棒控制问题

$$\begin{bmatrix} M_0 & U_1 \\ N_0 & V_1 \end{bmatrix} = \max_{(U_0, V_0) \in S_1(K_0; P_0)} \begin{bmatrix} M_0 & U_0 \\ N_0 & V_0 \end{bmatrix},$$

可通过以下标准的 H 四块问题求解:

$$\max_{(U_0, V_0) \in S_1(K_0; P_0)} \begin{bmatrix} M_0 & U_0 \\ N_0 & V_0 \end{bmatrix} = \left[\min_{X \in RH} \left\| \begin{bmatrix} R_1 - X & R_2 \\ & R_3 & R_4 \end{bmatrix} \right\| \right]^{-1}.$$

其中 X 和 $R_i (i = 1, 2, 3, 4)$ 的定义参见文献[5].

定理 2 最优鲁棒控制问题(13)和(14)可由下式求解:

$$\begin{aligned} \max_{(U_0, V_0) \in S_2(K_0; P_0)} \begin{bmatrix} M_0 & U_0 \\ N_0 & V_0 \end{bmatrix} &= \left[\min_{X \in RH} \left\| \begin{bmatrix} R_1 - X & R_2 \\ & R_3 & R_4 \end{bmatrix} \right\| \right]^{-1}, \\ \max_{(U_0, V_0) \in S_2(K_0; P_0)} \left\| \begin{bmatrix} M_0 & -U_0 \\ -N_0 & V_0 \end{bmatrix} \right\| &= \\ \min_{X \in RH} \left\| \begin{bmatrix} R_1 - X & -R_2 \\ -R_3 & R_4 \end{bmatrix} \right\|^{-1}. \end{aligned}$$

证明与引理 4 的证明类似,此略.

5 结 论

本文研究对象和控制器同时存在互质因子摄动时的鲁棒解耦控制问题,这是一个具有重要意义但未被充分重视的问题.本文通过研究闭环系统的鲁棒稳定性以及鲁棒解耦性能,给出了系统鲁棒稳定且具有鲁棒解耦性能的充分条件,并将最优鲁棒控制设计问题转化为标准的 H 四块问题,为非脆弱控制的进一步研究奠定了基础.

参考文献(References)

[1] Keel L H, Bhattacharyya S P. Robust, fragile or

optimal[J]. IEEE Trans on Automatic Control, 1997, 42(8): 1098-1105.

[2] Yue D, Lam J. Non-fragile guaranteed cost control for uncertain descriptor systems with time-varying state and input delays[J]. Optimal Control Applications and Methods, 2005, 26(2): 85-105.

[3] Yang G H, Wang J L. Robust nonfragile Kalman filtering for uncertain linear systems with estimator gain uncertainty [J]. IEEE Trans on Automatic Control, 2001, 46(2): 343-348.

[4] Boukas E K. Nonfragile controller design for linear Markovian jumping parameters systems [J]. J of Optimization Theory and Applications, 2004, 122(2): 241-255.

[5] 张高民, 贾英民. 基于互质因子摄动的反馈系统的鲁棒稳定性[J]. 自动化学报, 2002, 28(6): 974-976.

(Zhang G M, Jia Y M. Robust stability of feedback system with coprime factor perturbations [J]. Acta Automatica Sinica, 2002, 28(6): 974-976.)

[6] Loh A P, Quek C K. On the use of Hadamard weightings in the partial decoupling of H based designs [J]. Systems and Control Letters, 1995, 25(3): 193-198.

[7] 王松桂, 贾忠贞. 矩阵论中不等式[M]. 合肥: 安徽教育出版社, 1994.

(Wang S G, Jia Z Z. Inequalities in matrix theory[M]. Hefei: Anhui Education Press, 1994.)

[8] 姜长生, 孙隆和, 吴庆宪, 等. 系统理论与鲁棒控制[M]. 北京: 航空工业出版社, 1998.

(Jiang C S, Sun L H, Wu Q X, et al. System theory and robust control [M]. Beijing: Aviation Industry Press, 1998.)

(上接第 572 页)

[9] Pearl J. Statistics and causal inference: A review [J]. Sociedad de Estadística e Investigación Operativa Test, 2003, 12(2): 281-345.

[10] 王双成, 苑森森. 具有丢失数据的可分解马尔科夫网络结构学习[J]. 计算机学报, 2004, 27(9): 1221-1228.

(Wang S C, Yuan S M. Learning decomposable Markov network structure with missing data [J]. Chinese J of Computers, 2004, 27(9): 1221-1228.)

[11] 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): 1030-1041.

(Wang S C, Yuan S M. Research on learning Bayesian networks structure with missing data [J]. J of Software, 2004, 15(7): 1030-1041.)

[12] Chickering D M. Learning equivalence classes of Bayesian network structures [J]. Machine Learning, 2002, 2(3): 445-498.