

文章编号: 1001-0920(2006)05-0487-05

概率SDG模型及故障分析推理方法

杨帆, 萧德云

(清华大学自动化系, 北京 100084)

摘要: 符号有向图(SDG)是用来表示大规模复杂系统中变量之间因果影响关系的一种重要工具,但其存在一些不易克服的缺点。为此,首先提出一种新的模型——概率SDG模型,用条件概率描述故障之间的传递关系;然后在概率SDG模型的框架下,提出一种故障分析诊断的推理方法,即利用图消去算法和连接树算法进行贝叶斯推理,并计算出故障概率。最后以65 t/h锅炉系统为例,研究建立其概率SDG模型,并在此基础上验证了上述模型和推理方法的有效性。

关键词: 符号有向图; 安全评价; 故障分析; 贝叶斯网络

中图分类号: TP206 **文献标识码:** A

Probabilistic SDG Model and Approach to Inference for Fault Analysis

YANG Fan, XIAO De-yun

(Department of Automation, Tsinghua University, Beijing 100084, China Correspondent: XIAO De-yun, E-mail: xiaody@mails.tsinghua.edu.cn)

Abstract: Signed directed graph (SDG) model is a significant tool to express cause-effect relationships between variables in large-scale complex systems, but it also has some disadvantages that are difficult to overcome. A new kind of model, probabilistic SDG model, is proposed to describe the dependence relationships with conditional probabilities. In the framework of probabilistic SDG model, inference approach is presented, which implements Bayesian inference with elimination algorithm and junction tree algorithm to calculate the fault probability. Finally, a probabilistic SDG model is established for a typical instance of 65 t/h boiler system, which proves the validation of the model and inference approach.

Key words: Signed directed graph; Hazard assessment; Fault analysis; Bayesian network

1 引言

符号有向图(SDG)由节点和支路组成,节点表示系统变量(“+”、“-”或“0”),支路表示变量之间的因果影响关系(“+”或“-”)。这样的模型具有直观、清晰的特点,是大规模复杂系统的一种重要描述方式^[1]。基于这种模型的推理可用于大型企业的故障分析^[2]和安全评价^[3],完备性很好。然而,定性SDG模型也有很突出的缺点——冗余、虚假解多而且故障只限于变量值偏差的故障,这是定性SDG模型自身的缺陷。为了更好地利用各种信息,并引入随

机因素,变决定性关系为条件概率,于是形成了贝叶斯网络。为了表示更多类型的故障,应把变量节点与故障节点区分开来,用故障节点之间的支路表示故障之间的影响传递关系。

根据上述思路,本文建立了概率SDG模型,并探讨了其故障推理问题。首先给出概率SDG模型的基本概念,并与定性SDG模型进行对比;然后提出一种基于概率SDG模型的故障推理方法——消去算法和连接树算法,并与定性SDG推理方法作了对比。Matlab工具箱BN T是可以利用的软件工具;最

收稿日期: 2005-03-31; 修回日期: 2005-06-07

基金项目: 国家863计划项目(2003AA412310)。

作者简介: 杨帆(1980—),男,北京人,博士生,从事安全评估、故障诊断研究; 萧德云(1945—),男,福建仙游人,教授,博士生导师,从事辨识建模、故障诊断等研究。

后针对一个典型实例——65 t/h 锅炉, 建立其概率 SDG 模型, 并通过推理验证了上述算法的有效性

2 概率 SDG 模型

2.1 基本概念

概率 SDG 模型是在图上赋予概率分布得到的模型, 如用 4 个变量 x_1, x_2, x_3, x_4 表示系统的状态和故障模式, 则系统的状态可用联合分布 $p(x_1, x_2, x_3, x_4)$ 表示 根据 Bayes 公式进行边缘化展开, 得到

$$p(x_1, x_2, x_3, x_4) = p(x_1)p(x_2|x_1)p(x_3|x_1, x_2)p(x_4|x_1, x_2, x_3). \quad (1)$$

如果在 4 个变量之间又存在一定的依存关系, 即条件概率关系, 那么用变量作为节点, 条件概率关系作为支路, 可以构成有向图, 也称贝叶斯网, 如图 1 所示

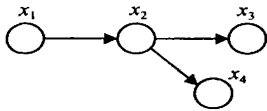


图 1 有向图示例

依据图 1 的变量依存关系, 3 条支路分别用条件概率可以表示成 $p(x_2|x_1), p(x_3|x_2)$ 和 $p(x_4|x_2)$, 因此式 (1) 可简化为

$$p(x_1, x_2, x_3, x_4) = p(x_1)p(x_2|x_1)p(x_3|x_2)p(x_4|x_2). \quad (2)$$

可见, 利用了变量之间的依存关系, 可以简化对系统状态的描述 仍以上例为例, 假设每个变量的取值可能性有 r 个, 则计算复杂度由 $O(r^4)$ 减为 $O(r^2)$, 而且这种简化在图的结构稀疏时更为突出

由变量节点和故障节点组成的 SDG 模型称为概率 SDG 模型, 定义如下:

定义 1 二元组 (G, P) 称为概率 SDG, 若:

- 1) 有向无环图 $G = (V, E)$, 其节点集合 $V = F \cup A$, 其中: F 表示故障类型和模式, A : 表示过程变量, 有向边集合 $E \subseteq (F \times F) \cup (F \times A) \cup (A \times V)$ 表示故障之间或变量与故障之间的条件概率关系
- 2) F 中每个元素的取值为“1”或“0”, “1”表示故障发生, “0”表示故障未发生(当然, 也可用多个取值表示多个故障状态); A 中每个元素的取值有 3 种情况(“+”、“-”、“0”), 与定性 SDG 的含义相同
- 3) P 为节点的概率分布 $\forall a \in V$, 记 π_a 为 a 在图 G 中的父节点集合, x_a 为 a 所对应的过程变量, 则图 G 表示的系统状态可用

$$p(x_1, \dots, x_n | \pi) = \prod_{a \in V} p(x_a | \pi_a)$$

来表示

概率 SDG 模型是故障和变量构成的有向无环图, 这是与定性 SDG 模型所不同的 对于存在环路

的情况, 需要具体分析其故障传递情况, 将其转化为概率 SDG 模型

在大规模复杂系统中, 各故障之间的关系可用概率 SDG 来描述 系统的故障类型包括操作错误、设备失灵、参数值偏离等许多种 在概率 SDG 中, 将故障表示成节点, 如阀门节点表示阀门操作状态, 设备失灵表现为一系列现象, 本身设立故障模式节点 例如, 一个管道, 其流量用流量计检测并显示出来, 安装一个自动控制的阀门控制流量, 这样一个简单的过程就可以用概率 SDG 模型表示出来 首先是变量节点——流量, 设定一个正常的流量范围, 高于上限为“+”, 低于下限为“-”, 在范围内为“0”, 这是一个可测节点; 流量值偏差为相应的故障; 流量由阀门的开度决定, 因为对阀门的误操作(如开大或关小)会导致流量的异常, 故设置一个故障节点来表示这一操作故障; 阀门是由控制器控制的, 控制器的故障会导致阀门操作故障, 故在阀门操作故障节点的前方设定控制器故障节点; 此外, 管道泄漏、阀门阻塞等原因都会带来流量的异常, 每个故障原因都可以用一个故障节点来表示, 这些故障节点都有两个状态: 正常或故障 在上述故障设定下的概率 SDG 模型如图 2 所示

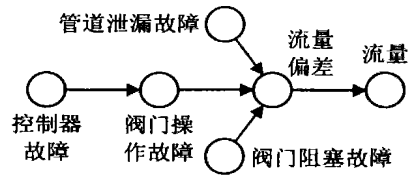


图 2 阀门控制流量过程的概率 SDG 模型

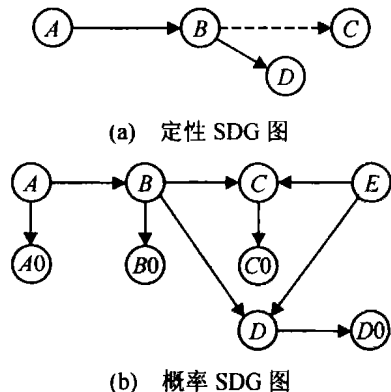


图 3 定性 SDG 图和对应的概率 SDG 图

下面以一个例子说明定性 SDG 是如何转化为概率 SDG 的 定性 SDG 如图 3(a) 所示, 共有 4 个节点 A, B, C, D , 表示 4 个变量, 箭头的实虚表示了变量之间影响关系的正反 概率 SDG 如图 3(b) 所示, 其中的节点除了定性 SDG 中的变量节点(这里用 A_0, B_0, C_0, D_0 表示)以外, 还有表示故障的节点

故障节点分两类: 一类与单一的变量节点相对应, 如 A, B, C, D 表示变量值的偏离误差, 由于变量也可表示阀门等人为操作值, 所以它们表示了操作错误或参数值偏差 比如 A 的值偏高会带来 B 值偏高、 C 值偏低、 D 值偏高 在实际系统中, 每个变量值的偏离往往对应若干个原因, 本例中只表示为一个, 即 A 对应 A_0, B 对应 B_0 , 依此类推 另一类与多个故障节点相连, 表示设备失灵等复杂的故障模式, 如故障 E 发生时, C 和 D 值偏高, 故从节点 E 到 C 和 D 各引一条支路 当故障影响到某个变量并继续沿故障传播关系传递下去, 例如某故障影响变量 B , 并进而影响 C 和 D , 则将支路终点置于 B ; 但如果此故障仅影响到变量 B 的值而不再传播, 则应将支路终点置于 B_0 这是概率 SDG 中将变量节点和变量值偏差故障节点分开的好处

图的结构出来后, 需赋予其条件概率分布才能表示故障之间的传递关系 事实上, 这些概率应该是通过先验知识以及对大量样本的学习得来的

2.2 与定性 SDG 模型的对比

概率 SDG 模型是定性 SDG 模型的推广, 它继承了定性 SDG 模型的许多概念和思路, 但也有许多不同之处, 表 1 对两者的不同作了总结

表 1 定性 SDG 模型与概率 SDG 模型的对比

	定性 SDG 模型	概率 SDG 模型
节点含义	变量	变量和故障
节点取值	“+”、“-”、“0”	有限个
支路关系	定性决定	条件概率
支路连接的节点	变量节点	故障节点、故障和变量节点之间
表示故障类型	变量值偏差	变量值偏差、操作错误、设备故障

可见, 概率 SDG 模型比定性 SDG 模型的适用面和灵活度大大增强, 其优势在推理方法上的体现见第 3 节

3 基于概率 SDG 模型的故障推理方法

3.1 图消去算法及连接树算法

基于概率 SDG 模型的故障推理是由观测到的样本计算各故障发生的概率 由于图模型表示的是条件概率, 所以实际上就是在图上的贝叶斯推理

假设概率 SDG 图的节点集合为 $V = E \cup F \cup H$, 其中: E 是证据节点, 即已知概率节点; F 是查询节点, 即待求概率节点; H 是隐藏节点, 即推理中不关心的节点 推理过程就是在 x_E 已知的条件下, 求 x_F 的条件概率

$$p(x_F | x_E) = \frac{p(x_E, x_F)}{p(x_E)}, \quad (3)$$

其中

$$p(x_E, x_F) = \sum_{x_H} p(x_E, x_F, x_H), \quad (4)$$

$$p(x_E) = \sum_{x_F} p(x_E, x_F). \quad (5)$$

变量消去与节点消去算法是图模型推理的基本方法^[4], 但它们只适用于查询节点是单个节点的情形, 如果想同时查询多个节点(这在故障推理中表现为考查多个故障的发生概率), 则必须反复计算, 这将带来大量的冗余 一个较成熟的算法——连接树(junction tree)算法^[4-6], 可以解决这些问题, 它是上述消去算法的推广

根据连接树算法一次计算可查询多个节点的特点, 将其与反向深度优先搜索相结合, 可完成故障源的搜索, 同时勾勒出故障传播的路径 算法如下:

```

BEGIN
  INTEGER i; COMMENT 全局变量
  PROCEDURE DFS(v, u); COMMENT v 是
    u 的父节点
    BEGIN
      NUMBER(v) := i := i + 1;
      COMMENT 为节点标号, 以记录搜索顺序
      计算 v 的父节点的概率; COMMENT
      用连接树算法 FOR v 存在概率大于阈值的
      父节点 w DO
        BEGIN
          IF w 没有被标号 THEN
            BEGIN
              DFS(w, v);
            END;
          END;
        END;
      END;
    END;
  i := 0;
  DFS(s, 0); COMMENT s 为任意一个状态异常
  的变量节点
END;
    
```

由于深度优先搜索是线性复杂度的, 因此算法的瓶颈在于: 在消去算法或 Junction tree 算法中, 算法的计算复杂度取决于最大簇 C_i , 即最大的完全子图, 若每个变量的取值可能性有 r 个, 则总的计算量为 $O(r^{\max |C_i|})$, 而决定簇的分解的是消去的序 使最大簇的规模最小, 这是个优化问题:

$$\min_I (\max |C_i|),$$

其中, 序 I 的选择有 $|V|!$ 种 这个问题已被证明是 NP 完全问题^[7], 因此, 只能利用启发式的方法或根据问题的特点寻找有效方法

建立概率 SDG 模型以及在其上应用此算法的完整步骤如下:



- 1) 建立定性SDG模型
- 2) 如果存在环路,则分析其故障与变量的影响关系,打开环路
- 3) 为每个变量添加变量值偏差故障节点和其他故障节点,并参考定性SDG,按照故障的传递关系连接这些故障节点
- 4) 按照确定性关系设置概率分布,即概率值只取0或1
- 5) 通过实验获取已知典型故障发生时的变量取值,作为证据,进行推理,求得各个故障的状态,用以判断模型的正确性
- 6) 收集各种情况下的变量值样本,进行学习,确定故障的概率分布
- 7) 实际应用中,将变量取值作为证据经过推理可得到各个故障的发生概率

3.2 与定性SDG推理方法的对比

SDG方法的核心是推理,定性SDG推理事实是在进行图的遍历搜索,该搜索是沿着相容通路进行的。推理的目的从图的角度可以概括为寻找强连通单元之间的序关系。如果是故障诊断,则只要找到极大强连通单元即可^[1]。为提高推理结果的准确性,算法往往包含两个步骤:首先通过反向推理寻找可能的故障源;然后从这些可能的故障源开始正向推理进行验证,如果模型引入了定量信息,最终得到的结果将是根据可能性的排序。

如果合并变量值偏差故障节点和变量节点,不考虑其他的故障,并将每个变量的概率都表示成0或1,则支路关系表示确定性传递关系,这时的概率SDG就是定性SDG。因此,定性推理可以视为概率推理的特例,赋予了条件概率之后的概率SDG图将具有更为丰富的涵义,这正是贝叶斯推理的意义所在,也是定性推理所不及的。

定性SDG推理由于是在相容通路的基础上进行搜索的,故在寻找故障源的同时就确定了故障传播通路;而概率SDG推理使用的是贝叶斯推理,虽然每个故障的概率都可求得,但故障通路还需要再作一次相容性判断,此判断以推理出来的故障概率为依据。显然,由于不可测节点的存在,定性SDG推理遇到不可测节点时只能假设,而概率SDG推理则对每个故障都可通过贝叶斯推理得到故障概率。依据故障概率的相容性分析尚待进一步研究,例如阈值选择等问题。

3.3 软件工具介绍

Matlab软件在数学分析和工程计算方面发挥了重要的作用, Kevin Murphy等近年开发了一个Matlab工具箱BNT,用于贝叶斯网的建模、推理与

学习,而概率SDG模型事实上是贝叶斯网络的一个特殊情况,所以可利用BNT开展概率SDG的研究。用BNT完成建模和推理的过程分为3步:

- 1) 通过邻接矩阵建立有向无环图;
- 2) 设定变量的概率分布;
- 3) 完成贝叶斯推理

由于BNT没有提供GUI,故只能用命令的方式执行操作。法国NSA-Rouen的学生编写了一个附加软件BNet Editor,实现了部分命令的图形化,但推理引擎目前只有Junction tree算法,且图形显示尚不友好。

4 65 t/h 锅炉的概率SDG模型和故障推理

选择65 t/h中压蒸汽锅炉系统作为实验对象,该系统在我国大型石油化工企业中应用广泛。本实验使用仿真软件PS^[8],工艺流程如图4所示。锅炉为双汽包结构,锅水自然循环产生蒸汽;入水经过热力除氧、省煤器换热等过程给到锅炉里;热量由高压瓦斯燃烧提供。流程中的重要变量用单回路控制,阀门可自动/手动调节。

该模型共有监测变量17个,分别是:锅炉上水流量FR-01,过热蒸汽输出流量FR-02,减温水流量FI-03,软化水流量FR-04,烟气流量FI-06,燃料油流量FR-07,至催化除氧水流量FI-08,炉膛压力PI-03,烟气出口压力PI-05,烟气含氧量AI-01,主蒸汽压力PIC-01,高压瓦斯压力PIC-02,液态烃压力PIC-03,除氧器压力PIC-04,上汽包水位LIC-01,除氧器水位LIC-02和过热蒸汽温度TIC-01。流程中的关键变量还有炉膛温度TI-07以及高、中、低压瓦斯流量和空气进气流量。该系统中还有调节阀C1-8;可调挡板或阀门ST1-6;HV11, 22~25, 29, 30, 32;V01, 02, 06, 09, 10, 18, 26, 27, 35~44, 46, 47, 50, 51, 54, 55和开关K01, 02, 03, 05。以上每个变量都构成一个变量节点,并建立与之相对应的变量之偏差故障节点(结尾加“p”)。变量之间的影响关系根据经验知识连成支路^[8,9]。

对于复杂的故障模式,引入5个常见故障:锅炉满水(F2)、锅炉缺水(F3)、锅炉灭火(F4)、电源中断(F5)和减温器坏(F6)。为每个故障建立一个故障节点,并将它们与受其影响的变量值偏差节点相连,构成了完整的概率SDG模型(如图5所示),并通过经验知识和对样本的学习赋予各个支路的条件概率分布。

通过BNT工具箱用连接树算法对上述模型进行故障推理,将检测到的变量样本作为证据节点,输入各个故障模式下的变量取值,即可推理得到变量值偏差故障和F2~F6的故障概率。例如,当仅有

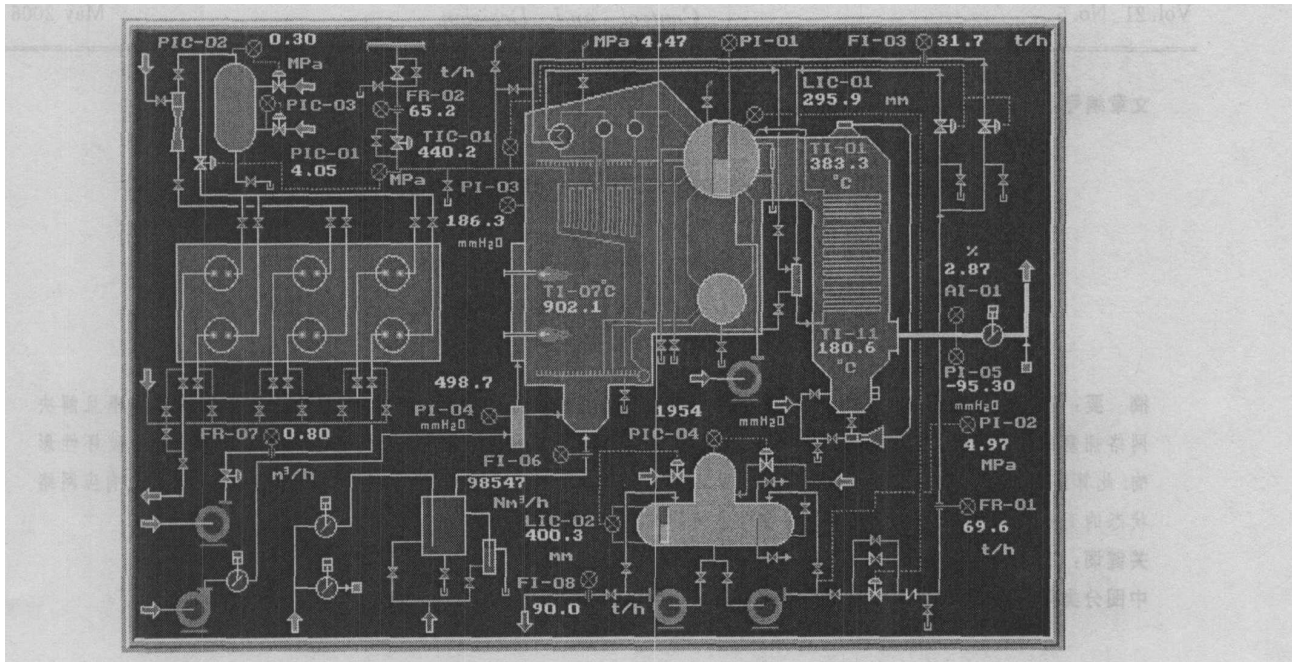


图4 65 t/h 锅炉的工艺流程^[8]

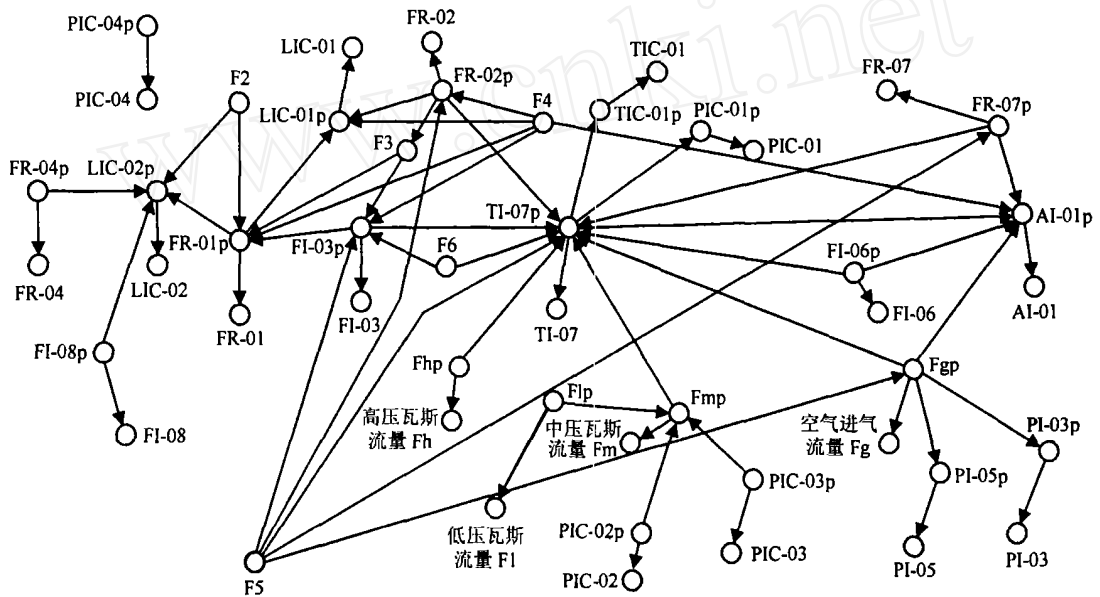


图5 65 t/h 锅炉的概率SDG 模型

FR-01 和 LIC-01 变量值偏差时,可以判定FR-01 的偏差为故障源,对应实际过程,则可判定是C3 故障;当仅有 TIC-01, F1-03 偏差时,可以判定 F6 故障发生,即减温器损坏

由于本例并没有在实际中大量实验,对故障的概率分布信息所知甚少,因此尚未体现随机因素引入的明显优势,但已能完成定性SDG 推理的所有功能,而且对于 F2~ F6 等复杂故障,基于定性SDG 的推理无法完成,而概率SDG 则能很好地胜任

5 结 论

SDG 模型是研究大规模复杂系统故障分析的

重要工具,但目前的研究主要是基于定性SDG 模型的,它还存在许多缺陷难以克服 本文提出了一种新的模型——概率SDG 模型,用节点明确表示了各种故障,用条件概率描述概率之间的传递关系,适于各种复杂故障的描述与传递关系研究 基于概率SDG 模型的故障分析实质上是个推理过程,而概率SDG 模型是贝叶斯网络的一种形式,所以可使用网络上的推理方法——消去算法和连接树算法完成推理计算,求得测量情况下的故障发生概率,判别故障模式 本文引入 Matlab 工具箱 BNT 完成推理,并对 65 t/h 锅炉建立了概率SDG 模型,验证了推理方法

(下转第 496 页)

参考文献(References)

- [1] Jain R. Congestion Control and Traffic Management in ATM Networks: Recent Advances and a Survey [J]. *Computer Networks and ISDN Systems*, 1996, 28(13): 1723-1738
- [2] 张孝林, 吴介一, 何小燕, 等. 基于速率的流量控制机制的研究[J]. *通信学报*, 2001, 22(5): 57-62
(Zhang X L, Wu J Y, He X Y, et al. Study of Rate Based Flow Control Mechanisms [J]. *J of China Institute of Communications*, 2001, 22(5): 57-62)
- [3] Kamolphiwong S, Karbowiak A E, Mehrpour H. Flow Control in ATM Networks: A Survey [J]. *Computer Communication*, 1998, 21(11): 951-968
- [4] Liping An, Nirwan Ansari, Ambalavanar Arulamalam. TCP/IP Traffic over ATM Networks with FMMRA ABR Flow and Congestion Control [J]. *Computer Networks and ISDN Systems*, 1998, 29(3): 2091-2102
- [5] ATM Forum. Traffic Management Specification Version 4.0 [S]. 1996
- [6] 姜培刚, 李春文, 吕景飞, 等. 时延高速通信网络的指数稳定跟踪控制[J]. *系统仿真学报*, 2004, 16(3): 546-550
(Jiang P G, Li C W, Lv J F, et al. Robust Tracking Stabilization for the Communication Networks with Multiple Time-delays [J]. *J of System Simulation*, 2004, 16(3): 546-550)
- [7] 沈伟, 胡立生, 邵惠鹤. 网络流量控制系统的设计——Kalman 算法[J]. *通信学报*, 2003, 4(24): 49-57.
(Shen W, Hu L S, Shao H H. The Design of Traffic Control System in Networks — Kalman Control Algorithm [J]. *J of China Institute of Communications*, 2003, 4(24): 49-57.)
- [8] 胡寿松. *自动控制原理* [M]. 第3版. 北京: 国防工业出版社, 1999.
(Hu S S. *Modern Control Theory* [M]. 3rd ed. Beijing: National Defence Industry Press, 1999.)
- [9] 张孝林, 吴介一, 朱正强, 等. ATM 网络中面向 ABR 服务的一种流量控制机制[J]. *电子学报*, 2002, 30(4): 503-507.
(Zhang X L, Wu J Y, Zhu Z Q, et al. A Flow Control Mechanism for ABR Service in ATM Networks [J]. *Acta Electronica Sinica*, 2002, 30(4): 503-507.)

(上接第 491 页)

参考文献(References)

- [1] Iri M, Aoki K, O'shima E, et al. An Algorithm for Diagnosis of System Failures in the Chemical Process [J]. *Computers and Chemical Engineering*, 1979, 3(1-4): 489-493
- [2] Chang C C, Yu C C. On-line Fault Diagnosis Using the Signed Directed Graph [J]. *Industrial and Engineering Chemistry Research*, 1990, 29(7): 1290-1299
- [3] Venkatasubramanian V, Zhao J, Viswanathan S. Intelligent System for HAZOP Analysis of Complex Process Plants [J]. *Computers and Chemical Engineering*, 2000, 24(9/10): 2291-2302
- [4] Cowell R G, Dawid A P, Lauritzen S L, et al. *Probabilistic Networks and Expert Systems* [M]. New York: Springer-Verlag, 1999
- [5] Pearl J. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference* [M]. San Mateo: Morgan Kaufmann, 1988
- [6] Jensen F V. *Bayesian Networks and Decision Graphs* [M]. New York: Springer, 2001
- [7] Cooper G F. The Computational Complexity of Probabilistic Inference Using Bayesian Belief Networks [J]. *Artificial Intelligence*, 1990, 42(2-3): 393-405
- [8] 吴重光. 用于操作分析的 65t/h 环保锅炉仿真建模[J]. *锅炉技术*, 2002, 33(9): 1-6
(Wu C G. Modeling and Simulation of 65t/h Steam Boiler for Operation Analysis [J]. *Boiler Technology*, 2002, 33(9): 1-6)
- [9] 张贝克. SDG(符号有向图)实时推理机制图形化软件平台及安全工程应用研究[D]. 北京: 北京化工大学, 2003
(Zhang B K. Study with SDG(Signed Directed Graph) on Its Real-time Deduction Mechanism, Graphic Software Platform and Application in Safety Engineering [D]. Beijing: Beijing University of Chemical Technology, 2003)