

基于贝叶斯网络的计算机经济信息 系统可靠性模型及其应用

钱争鸣, 管 睿

(厦门大学 经济学院, 福建 厦门 361005)

摘 要: 主要针对计算机经济信息系统由构件和连接件形成的可靠性模型及其应用进行了研究。根据现代贝叶斯网络方法具有推理精确、直观易理解和较强可操作性的特点, 经过对经济信息系统的实例分析可以得到系统可靠性计算概率值与系统诊断的分析结果。说明该模型可方便地用于经济信息系统的可靠性计算和系统可靠性诊断, 并可以在第三方构件和异构环境中进行可靠性分析。

关键词: 贝叶斯网络; 计算机经济信息系统; 可靠性模型

中图分类号: TP309.1 **文献标志码:** A **文章编号:** 1001-3695(2012)12-4651-03

doi:10.3969/j.issn.1001-3695.2012.12.063

Application of reliability model about computer business information system based on Bayesian network

QIAN Zheng-ming, GUAN Rui

(School of Economics, Xiamen University, Xiamen Fujian 361005, China)

Abstract: This paper proposed a new reliability model formed by the components and connectors and its application for the computer business information system. The model was based on Bayesian network which was accurate, clear and operable. It concluded the system reliability probability and diagnosis result through the case analyzing. This model applied a method more convenience in testing and diagnosing reliability of the business information system, and broadens the scope of the application in the heterogeneity component environment.

Key words: Bayesian network; computer business information system; reliability model

0 引言

可靠性是重要的质量指标,随着计算机信息系统开发和应用中出现的与可靠性相关的问题不断增加,直接推动了信息系统的可靠性研究与定量分析。经济信息系统设计与开发过程及其工程实践证明,无论拥有多么优秀的开发与编程人才,采用了多么好的体系结构和构建过程,进行了多么有效的管理,在经济信息系统开发的整个生命周期中,总会或多或少地出现一些可靠性问题,而问题发现得越早,解决起来就越容易,付出的代价也越小。从计算机信息系统可靠性概念出发,人们已先后建立起若干计算机信息系统可靠性模型,并提供了一些评估计算机信息系统可靠性的有效方法。

计算机信息系统可靠性的研究可追溯到 20 世纪 60 年代。1967 年 Hudson 首先提出以开发过程中产生差错开始,以差错修正而结束的可靠性分析方式。1973 年, Wolverson 基于对危险率同残留差错数和时间的乘积成比例的认识,提出将危险率视为随机变量,其分布函数含有一个随经历的故障数变化的参数,而构建一种新的计算机信息系统可靠性模型。1975 年, Musa 在总结前人工作的基础上提出可靠执行时间模型。20 世纪 80 年代之后,开始对计算机信息系统可靠性模型进行比较和选择研究。虽然已经取得一定的成绩,

但目前理论与应用之间仍存在很大距离,应用情况不尽如人意。主要原因是计算机信息系统可靠性模型的精度不高、适应性不好;对信息系统故障所作的假设固定不变,而计算机信息系统却受到各种随机因素的综合影响等方面的问题。通过比较分析,笔者发现已提出的计算机信息系统可靠性的估算模型大多数是围绕信息系统测试阶段的可靠性增长模型,它没有考虑模型的内部结构,近似于黑箱模型。然而,随着面向对象的系统设计和基于 Web 的计算机信息系统开发方法的广泛应用,特别是基于构件的计算机信息系统开发模式正在不断增长,经济信息系统构件可以是第三方构件、自主研发的构件或合约研发的构件,整个应用开发是在一个异构环境下。因此,应用现有的计算机信息系统可靠性评价模型以及相应的研究与分析方法,对现代计算机信息系统的可靠性分析,特别是基于构件的信息系统可靠性进行估算和应用分析是不合适的。

为此,本文构建一种新的基于构件与连接件可靠性评估方法的经济信息系统可靠性贝叶斯网络模型。该方法根据构件和连接件的可靠性概率,以及整个计算机经济信息系统的条件概率,推导出信息系统的可靠性,具有推理精确、直观易理解的特点,且具有较强的可操作性。此外,该模型还适用于异构环境下的计算机经济信息系统可靠性测试。

收稿日期: 2012-05-16; 修回日期: 2012-06-28

作者简介: 钱争鸣(1956-),男,江苏泰兴人,教授,博导,博士,主要研究方向为经济信息管理和统计学(zmqianxm@126.com);管睿(1983-),男,上海人,博士,主要研究方向为统计学。

1 贝叶斯网络

贝叶斯网络最早由 Pearl 于 1988 年提出, 又称为信念网络, 它使用概率理论来描述和处理由不同知识成份的条件相关而产生的不确定性, 适用于不确定性和概率性事物, 实质上是一种基于概率的不确定性推理网络。其用来表示变量集合连接概率的图形模型, 是一种对概率关系作有向图解描述, 提供了一种将知识进行直觉图解可视化的方法, 以及一种表示因果关系的方法。贝叶斯网络主要是概率理论与图论相结合的产物, 具有严格的概率论基础。而概率论是人工智能中处理不确定性的基础理论之一, 也被认为是数学基础最强的不确定性处理理论。

一个贝叶斯网络主要由贝叶斯网络结构和网络参数两部分构成, 分别对应问题领域的定性描述和定量描述。贝叶斯网络结构是一个有向无环图 (DAG), 由节点集合和有向边集合组成。节点集中的每个节点代表一个随机变量, 变量可以是任何问题的抽象, 用来代表研究与分析对象的现象、部件、状态或属性等, 具有一定的物理和实际意义。有向边表示变量之间的依赖或因果关系, 有向边的箭头代表因果关系影响的方向性 (由父节点指向子节点), 节点之间若无连接边, 表示节点所对应的变量之间是条件独立的。指向节点 X 的所有节点称为 X 的父节点。尽管从节点 X 指向节点 Y 的有向边被用来表示 X 引起了 Y 的变化, 但在贝叶斯网络中并不是对有向边的唯一解释。例如, Y 可能只与 X 有关联, 但是它不是由 X 引起的。因此, 虽然贝叶斯网络能表示因果关系, 但它们并不局限于表示因果关系。

贝叶斯网络的另一部分是网络参数 (概率参数), 构成了反映变量之间关联性的局部概率分布集, 常称之为条件概率表 (conditional probability table, CPT), 表中列出了每个节点相对于其父节点所有可能的条件概率。贝叶斯网络约定以节点 X_i 的父节点为条件, X_i 与任意非 X_i 子节点条件独立。概率值表示子节点与其父节点之间的关联强度或置信度, 没有父节点的节点概率为先验概率。贝叶斯网络结构是将数据实例抽象化的结果, 是对问题领域的一种宏观描述。而概率参数是对变量 (节点) 之间关联强度的精确表达, 属于定量描述的部分。

贝叶斯网络在有联系的节点之间建立连接弧, 则有 n 个节点的贝叶斯网络联合概率分布为

$$\text{Pro}(x_1, x_2, \dots, x_n) = \prod_{i=1}^n \text{Pro}(x_i | \text{Pro}_{a_i})$$

其中: Pro_{a_i} 为节点 x_i 的父节点集合。

所以贝叶斯网络是目前不确定性知识表达和推理领域中最有效的理论模型之一。在解决实际问题中, 它需要从不安全的或不确定的知识和信息中推理, 适用于不确定性和概率推理的知识表达和推理。它是一种基于网络结构的有向图解描述, 能进行双向并行推理, 并能综合先验信息和样本信息, 使得推理结果更为准确可信。

2 基于贝叶斯网络的经济信息系统可靠性模型

在整个计算机经济信息系统中, 构件是计算机信息系统的主要组成部分, 而构件与构件之间是通过连接件连接形成一个完整的计算机信息系统。可以把所有的构件看成一个子系统, 并假设这些构件之间都是相互独立的; 同样把连接件看成一个

子系统, 并假设这些连接件之间都是相互独立的。这样整个计算机信息系统可以看成由构件子系统与连接件子系统组成的综合系统。通过运用各构件和连接件的可靠性, 估算构件和连接件条件下构件子系统和连接件子系统的系统可靠性概率, 运用贝叶斯网络模型来精确推理出整个计算机信息系统的可靠性。例如, 当某个经济信息系统由 n 构件、 m 连接件组成, 其满足的贝叶斯网络模型如图 1 所示。

构件 c_i 的可靠性为 Pro_{c_i} , 连接件 l_j 的可靠性 Pro_{l_j} , 系统的可靠性为

$$\text{Pro}_s = \text{Pro}(s | c, l) \text{Pro}(c, l) = \text{Pro}(s | s, l) \text{Pro}(c) \text{Pro}(l) \quad (1)$$

因为构件子系统与连接件子系统在贝叶斯网络中满足条件独立关系, 所以有 $\text{Pro}(c, l) = \text{Pro}(c) \text{Pro}(l)$ 。同样的道理对于构件子系统与连接件子系统有

$$\text{Pro}(c) = \text{Pro}(c | c_1, \dots, c_i, \dots, c_n) \text{Pro}(c_1, \dots, c_i, \dots, c_n) = \text{Pro}(c | c_1, \dots, c_i, \dots, c_n) \prod_{i=1}^n \text{Pro}(c_i) \quad (2)$$

$$\text{Pro}(l) = \text{Pro}(l | l_1, \dots, l_j, \dots, l_m) \text{Pro}(l_1, \dots, l_j, \dots, l_m) = \text{Pro}(l | l_1, \dots, l_j, \dots, l_m) \prod_{j=1}^m \text{Pro}(l_j) \quad (3)$$

因此, 把式 (2) (3) 代入式 (1), 可以得到整个计算机信息系统的可靠性为

$$\text{Pro}_s = \text{Pro}(s | c, l) \text{Pro}(c | c_1, \dots, c_i, \dots, c_n) \prod_{i=1}^n \text{Pro}(c_i) \text{Pro}(l | l_1, \dots, l_j, \dots, l_m) \prod_{j=1}^m \text{Pro}(l_j) \quad (4)$$

其中: $\text{Pro}(c_i)$ 表示的是计算机信息系统中第 i 个构件的可靠性; $\text{Pro}(l_j)$ 代表的是计算机信息系统中第 j 个连接件的可靠性; $\text{Pro}(c | c_1, \dots, c_i, \dots, c_n)$ 是在各个构件条件下, 构件子系统的可靠性; $\text{Pro}(l | l_1, \dots, l_j, \dots, l_m)$ 是在各个连接件的条件下, 连接件子系统的可靠性; $\text{Pro}(s | c, l)$ 是在构件子系统和连接件子系统下, 整个计算机信息系统的可靠性。

3 实例研究

贝叶斯网络的构建是一个复杂的过程。计算机信息系统可靠性分析的贝叶斯网络所需信息来自于多种渠道中汇集的各种构件与相应的连接件资料, 诸如设备手册、计算机信息系统设计与开发过程、测试过程、维护过程以及专家经验等。为了减少复杂性, 对信息系统可靠性分析的过程, 首先将计算机信息系统可靠性分析分为各个相互独立且完全包含的类别 (各种类别至少应该具有可以区分的界限); 然后对各个类别分别构建贝叶斯网络模型。需要注意的是计算机信息系统可靠性模型只在发生故障时启动, 因此无须对设备正常状态建模。通常计算机信息系统可靠性由某个或某些不同的原因造成, 这些原因又可能由某个或几个更低层次的原因造成。建立起网络的节点关系后, 还需要进行概率估计。具体方法是假设在某个原因出现的情况下, 估计该原因的各个节点的条件概率, 通过这种局部化概率估计的方法可以大大提高效率。

经济信息系统可靠性的贝叶斯网络模型有整体系统可靠性计算与系统诊断功能两大主要功能。

a) 系统可靠性计算。不失一般性, 本文将贝叶斯网络模型运用于一个实例, 该实例有四个构件和三个连接件, 基于贝叶斯网络的计算机信息系统可靠性模型如图 2 所示。其构件子系统的各构件的可靠性分别为 $\text{Pro}(C_1) = 0.933$, $\text{Pro}(C_2) = 0.983$, $\text{Pro}(C_3) = 0.908$, $\text{Pro}(C_4) = 0.992$, 连接件子系统的可靠性分别为 $\text{Pro}(l_1) = 0.92$, $\text{Pro}(l_2) = 0.967$, $\text{Pro}(l_3) = 0.95$ 。

构件子系统 C 在各构件 C_1, C_2, C_3, C_4 条件下的可靠性即节点 C 的条件概率如表 1 所示。

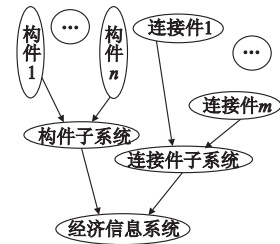


图1 构件子系统和连接件子系统

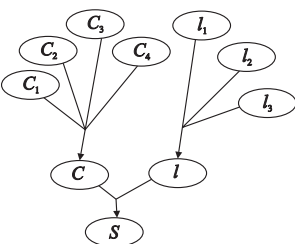


图2 贝叶斯网络

表 1 C_1, C_2, C_3, C_4 条件下可靠性条件概率

C_1	C_2	C_3	C_4	C	
				Yes	No
Yes	Yes	Yes	Yes	1.0	0.0
			No	0.681	0.319
		No	Yes	0.706	0.294
			No	0.48079	0.51921
	No	Yes	Yes	0.765	0.235
			No	0.52096	0.47903
		No	Yes	0.54009	0.45991
			No	0.3678	0.6322
No	Yes	Yes	Yes	0.782	0.218
			No	0.53254	0.46746
		No	Yes	0.55209	0.44791
			No	0.37598	0.62402
	No	Yes	Yes	0.59823	0.40177
			No	0.4074	0.5926
		No	Yes	0.42235	0.57765
			No	0.28762	0.71238

连接件子系统 L 在各连接件 L_1, L_2, L_3 的条件下的可靠性即节点 L 的条件概率如表 2 所示。

表 2 L_1, L_2, L_3 条件下可靠性条件概率

L_1	L_2	L_3	L	
			Yes	No
Yes	Yes	Yes	1.0	0.0
		No	0.815	0.185
	No	Yes	0.739	0.261
		No	0.60229	0.39772
No	Yes	Yes	0.866	0.134
		No	0.70579	0.29421
	No	Yes	0.63997	0.36003
		No	0.52158	0.47842

最后估算构件子系统和连接件子系统条件下的系统可靠性如表 3 所示。

表 3 构件子系统和连接件子系统可靠性概率

C	L	S	
		Yes	No
Yes	Yes	1.0	0.0
	No	0.706	0.294
No	Yes	0.227	0.773
	No	0.0	1.0

通过贝叶斯网络推理的原理,可以计算出构件系统的可靠性为 $\text{Pro}(c) = 0.9525$, 连接件子系统的可靠性为 $\text{Pro}(l) = 0.9717$, 整个系统的可靠性为 $\text{Pro}(s) = 0.9550$ 。

b) 系统诊断。诊断功能是假设计算机信息系统出现了故障,那么必须诊断出系统产生故障的原因是什么,从而进行修正。就上述实例而言,如果系统出现了异常状况,也就是由故障引起的不可靠性,那么可以利用贝叶斯模型的推理功能得出可靠性概率的计算值(表 4)。

表 4 贝叶斯网络算法可靠性概率

mode name	state 1(Yes)	state 0(No)
C	0.176 3	0.823 7
C_1	0.707 1	0.292 9
C_2	0.918 6	0.081 4
C_3	0.495 8	0.504 2
C_4	0.950 5	0.049 5
L	0.793 8	0.206 2
L_1	0.857 4	0.142 6
L_2	0.914 2	0.085 8
L_3	0.894 3	0.105 7
S	0.000 0	1.000 0

当系统 S 为不正常时,通过贝叶斯网络算法的推理,可以知道,构件子系统不可靠的概率为 $\text{Pro}(c = \text{No}) = 0.8237$, 而连接件子系统不可靠的概率为 $\text{Pro}(l = \text{No}) = 0.2062$, 可见最可能的异常原因在构件系统中。而构件 C_1, C_2, C_3, C_4 的可能性又分别为

$$\text{Pro}(C_1 = \text{No}) = 0.2929, \text{Pro}(C_2 = \text{No}) = 0.0814$$

$$\text{Pro}(C_3 = \text{No}) = 0.5042, \text{Pro}(C_4 = \text{No}) = 0.0495$$

可见,问题最可能出现在构件 C_3 。说明基于贝叶斯网络的计算机信息系统可靠性模型具有一个良好的诊断功能,能够帮助经济信息系统管理者找出系统不可靠时的原因所在。

4 结束语

贝叶斯网络将有向无环图与概率理论有机结合,不但可以将人类所拥有的因果知识直接用有向图自然直观地表示出来,还可以将统计数据以条件概率的形式融入模型。因此,贝叶斯网络能克服框架、语义网络等模型仅能表达、处理定量信息的弱点和神经网络等方法不够直观的缺点。本文提出的可靠性计算模型与其他计算机信息系统可靠性分析和评估方法相比具有以下优点:a)以前的系统可靠性分析只侧重于内部测试,贝叶斯网络可靠性模型分析弥补了其他模型方法对系统中连接件欠考虑的不足;b)该模型在实际运用中,具有良好的可操作性;c)贝叶斯网络具有良好的诊断功能,一旦发现系统不稳定,就可以通过诊断推理算法查出经济信息系统不可靠的原因所在,从而便于处理和解决问题;d)该模型可以运用于大型计算机信息系统,如第三方构件、计算机信息系统外包等异构环境中的可靠性测试,并能够顺利找出大系统不稳定的责任所在。因此,该模型拓宽了信息系统可靠性分析的应用范围。

参考文献:

[1] MUSA J D. A theory of software reliability and its application[J]. IEEE Trans on Software Engineering,1975,9(1):312-327.

[2] PEARL J. Probabilistic reasoning in intelligent systems; networks of plausible inference[M]. San Francisco: Morgan Kaufmann Publishers Inc,1988.

[3] PEARL J. Fusion, propagation and structuring in belief network[J]. Artificial Intelligence,1986,29(3):241-288.

[4] REUSSNER R H, SCHMIDT H W, POEMNOMO I H. Reliability prediction for component-based software architectures[J]. Journal of Systems and Software,2003,66(3):241-252.

[5] HUANG C Y, KUO S Y, LYU M R. An assessment of testing-effort dependent software reliability growth models[J]. IEEE Trans on Reliability,2007,56(2):198-211.

[6] MARQUEZ D, NEIL M, FENTON N. Improved reliability modeling using Bayesian networks and dynamic discretization[J]. Reliability Engineering & System Safety,2010,95(4):412-425.

[7] TAHERI S M, ZAREI R. Bayesian system reliability assessment under the vague environment[J]. Applied Soft Computing,2011,11(2):1614-1622.