

# 贝叶斯网络在软件项目风险评估中的应用

唐爱国<sup>1</sup>,王如龙<sup>2</sup>,胡春华<sup>1</sup>

TANG Ai-guo<sup>1</sup>,WANG Ru-long<sup>2</sup>,HU Chun-hua<sup>1</sup>

1.湖南商学院 计算机与电子工程学院,长沙 410205

2.湖南大学 软件学院,长沙 410082

1.College of Computer and Electronic Engineering,Hunan University of Commerce,Changsha 410205,China

2.College of Software,Hunan University,Changsha 410082,China

E-mail:tangaiquo\_2005@sina.com.cn

TANG Ai-guo,WANG Ru-long,HU Chun-hua.Application of Bayesian networks in software project risk assessment. Computer Engineering and Applications,2010,46(7):62-65.

**Abstract:** In the early phase of software project life cycle or the research and development process of innovative project,the available case data are few or incomplete,and project risk mainly experiences the subjective assessment of experts which brings great difficulties for the objective measure of risk.A software project risk assessment method based on Bayesian network is provided,which is not only capable of measuring the risk equivalent of the risk impact,but also able to measure out the combined effect of a variety of risks on the consequences of a certain risk or an individual risk on the overall consequences,thus increasing the prediction and response capabilities of the software project risk,and thus providing a new way that can effectively reduce the risk probability and improve the success rate of software development.

**Key words:** bayesian networks;software project;risk management;risk assessment

**摘要:**在软件项目生存周期早期或创新型项目的研发过程中,可用的案例数据很少或很不完整,项目风险多由专家经验进行主观评估,给风险的客观度量带来了很大的困难。提出了一种基于贝叶斯网络的软件项目风险评估方法,不仅可度量风险影响程度的风险当量,还能度量出多种风险对某种风险后果的组合影响以及单个风险对整体后果的综合影响,从而增强了软件项目风险的预测和应变能力,为有效地降低风险发生概率、提高软件开发成功率提供了一种新的途径。

**关键词:** 贝叶斯网络;软件项目;风险管理;风险评估

**DOI:**10.3778/j.issn.1002-8331.2010.07.019 **文章编号:**1002-8331(2010)07-0062-04 **文献标识码:**A **中图分类号:**TP311

## 1 引言

所谓“风险”,是损失的不确定性,是在给定情况下,一定时期内可能发生各种结果间的差异。与任何其他工程项目一样,软件项目的开发也存在各种风险,有些风险甚至是灾难性的<sup>[1]</sup>。软件项目风险是指在软件开发过程中遇到的费用、进度与质量等方面的问题以及这些问题对软件项目的影响。软件项目风险管理有利于降低软件项目开发风险,提高软件项目的成功率。软件项目风险管理过程分为风险评估和风险控制两大部分,其中风险评估分为风险识别、风险分析、风险学习三个过程,风险控制则包括风险应对和风险监控两个过程<sup>[2]</sup>。风险评估是风险管理的核心和基础,直接影响风险管理的其他过程,甚至影响到项目的成败。软件项目风险管理过程如图1所示。

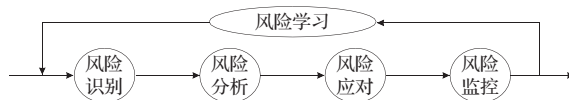


图1 软件项目风险管理过程图

近年来,软件项目风险管理得到了广泛的重视,但在软件开发项目中进行完整的规范化的风险评估还很不足,主要表现在软件项目风险管理的研究和实践起步较晚,风险管理停留在概念层面,缺乏具体管理制度和方法的支持;软件企业主要依靠项目管理人员的个人经验,实行以定性分析为主的风险管理,风险管理水平参差不齐,具有一定的随意性;由于在风险管理过程中缺乏定量分析方法,因此有关项目资料和数据收集具有盲目性,从而导致该领域的知识不能得到有效的积累,影

**基金项目:**国家科技支撑计划项目(the National Science and Technology Support Plan Project of China under Grant No.2006BAF01A13);湖南省自然科学基金(the Natural Science Foundation of Hunan Province of China under Grant No.09JJ4030)。

**作者简介:**唐爱国(1976-),男,讲师,研究方向:软件工程、IT项目管理;王如龙(1954-),男,教授,研究生导师,湖南省计算技术研究所研究员,享受政府特殊津贴专家,主要研究方向:企业信息化、软件工程、IT项目管理;胡春华(1973-),男,工学博士,副教授,研究方向:电子商务、Web服务工作流和跨企业协作。

**收稿日期:**2009-09-23 **修回日期:**2009-11-30

响了软件项目风险管理水平的提高。

## 2 风险评估研究现状

### 2.1 常见的风险评估方法

依据项目所能提供的经验数据的多少及信息详细程度,风险评估技术可分为定性与定量两种。定性分析主要是依靠专家确定项目的风险程度,专家经验决定了整个项目的综合考虑,主观性强;定量分析包括 AHP、模糊综合评判以及人工智能等方法。AHP 方法将复杂问题简单化,但比较判断矩阵的数据需求量大;模糊综合评判方法可同时针对多个目标,适用于对复杂项目的评估,但指标权重采用精确值,与实际情况有出入;人工智能方法则包括规则库、决策树和人工神经网络等方法,在一定程度上解决了项目评估的复杂性问题,但其有效性是以大量的训练样本为基础,影响了这些方法在项目评估中的应用。

近年来,软件项目风险评估受到了广泛关注,已有大量的研究成果。文献[1]将贝叶斯网络应用于软件项目风险评估,将软件项目风险管理划分为风险识别、风险分析、风险发生概率评估、风险影响评估及风险预防与控制等五个子模型;文献[3]对风险与风险管理的研究进行了重新评估,指出风险管理的研究滞后于实践的需要,并对未来的风险与风险管理的研究和实践方向进行了讨论;文献[4]提出了一种融合模糊理论以及专家信任度的风险评估模型,侧重以专家经验评判风险发生的概率。文献[5]针对软件项目的特点和软件项目风险定量评估任务,提出了基于面向对象贝叶斯网络的风险评估方法,并通过分析软件项目生命周期中的具体风险与风险因素之间的因果关系,建立面向对象贝叶斯网络拓扑结构,由专家经验确定网络中的概率参数,定量估计风险的发生概率。

通过真实的软件开发项目,根据已有案例启动风险案例学习机制,结合专家经验生成本次评估贝叶斯网络,以贝叶斯概率理论为基础,计算各风险节点发生的概率以及各风险因素对后果事件的综合影响,引入风险权重度量多种风险的组合影响。实践表明该方法为软件项目风险评估提供了一种有效的新途径。

### 2.2 贝叶斯网络方法研究软件项目风险评估的优点

贝叶斯网络(Bayesian Networks, BN)是图论与概率论的结合,为变量间概率关系的图形化描述提供了一种将知识直观的图解可视化的方法,同时又是一种概率推理技术,使用概率理论来处理在描述不同知识成分之间的因条件相关而产生的不确定性<sup>[6]</sup>。用贝叶斯网络作为软件项目风险评估工具,其优点主要体现在以下三个方面:

第一,贝叶斯网络的推理以贝叶斯概率理论为基础,具有成熟的概率推理算法和开发软件,为风险预测的贝叶斯模型建造和推理提供快捷的工具,加速了风险预测的有效性。

第二,贝叶斯网络将概率理论与有向无环图有机结合,既可以将因果知识直接用有向图直观地表示出来,也可以将统计数据以条件概率的形式融入模型,还能使人类的先验知识和经验数据无缝地结合。

第三,贝叶斯网络用图形化表示随机变量间的联合概率,能够处理各种不确定性信息,网络中没有确定的输入或输出节点,节点之间是相互影响的,更适合表达软件项目风险之间的关联关系。

## 3 贝叶斯网络的构建原理与方法

### 3.1 贝叶斯网络的形式化分析

根据对贝叶斯网络的描述,对于风险分析网络中的一组变量  $X=\{X_1, X_2, \dots, X_n\}$ ,将风险分析网络模型  $B_m$  定义为一个二元组:

$$B_m = \langle B_s, B_p \rangle \quad (1)$$

其中,  $B_s$  表示风险分析网络的结构,是一个具有  $n$  个节点的有向无环图,图中的每个节点代表一个变量  $X_i$ ,节点的状态对应于变量的值,图中的有向边表示了节点(变量)之间的条件(因果)依赖关系;  $B_p$  表示风险分析网络的条件概率分布表(CPT)集合,每个变量  $X_i$  的条件概率分布为  $P(X_i | Pa(X_i))$ ,其中  $Pa(X_i)$  为  $X_i$  的父结点的集合<sup>[7]</sup>。

风险分析网络结构  $B_s$  可细化为:

$$B_s = \langle N_r, N_f, R \rangle \quad (2)$$

其中,  $N_r$  表示风险分析网络中的风险节点,  $N_r \in X$ ,用矩形及矩形内的风险名来表示,每个风险对应唯一的风险 ID;  $N_f$  表示风险分析网络中的风险因子节点,  $N_f \in X$ ,用椭圆及椭圆内的风险因子名来表示,每个风险因子对应唯一的风险因子 ID;  $R$  表示风险分析网络中的关系,关系定义为一个二元序列对集合,用连接两个节点的箭头表示,箭头的起始端表示原因,终端表示结果。

风险分析网络提供了对域的完整描述,联合分布中的一般条目是对每个变量赋予一个特定值下的合取  $P$ 。可使用  $B_p$  即风险网络节点的 CPT 将分析网络中的概率信息计算出来。

用符号  $P(X_1, X_2, \dots, X_n)$  作为概率的简化表示,则有:

$$P(X_1, X_2, \dots, X_n) = \prod_{i=1}^n P(X_i | Pa(X_i)) \quad (3)$$

联合概率分布中的每个条目都可表示为 CPT 中适当元素的乘积。因此, CPT 提供了联合概率分布的一种分解表示方法。

### 3.2 贝叶斯网络的构建原理

给定一个联合概率分布  $P(X_1, X_2, \dots, X_n)$  以及变量的一个排序  $d$ ,将  $X_1$  作为根结点开始,并赋予  $X_1$  以先验概率分布  $P(X_1)$ 。然后用结点表示  $X_2$ ,如果  $X_2$  与  $X_1$  有关,从  $X_1$  到  $X_2$  建立联结,并用  $P(X_2 | X_1)$  表示联结强度。如果  $X_2$  与  $X_1$  无关,则赋予  $X_2$  以先验概率分布  $P(X_2)$ 。在第  $i$  级从  $X_i$  的父结点集合

$\prod_{X_i} \left[ \prod_{X_i} \subseteq \{X_1, X_2, \dots, X_{i-1}\} \right]$ ,画一组方向线联结到  $X_i$ ,并用  $P \left[ X_i | \prod_{X_i} \right]$  条件概率定量表示,结果可以得到一个有向无环图,可用于表示  $P(X_1, X_2, \dots, X_n)$  中所体现的许多独立关系,该图称作贝叶斯网络<sup>[8]</sup>。

反过来,  $P \left[ X_i | \prod_{X_i} \right]$  包含有重构原始分布函数所必需的所

$$\begin{aligned} \text{有信息,在排序 } d \text{ 下,有如下关系:} \\ P(X_1, X_2, \dots, X_n) = P(X_n | X_{n-1}, X_{n-2}, \dots, X_1) \cdot \\ P(X_{n-1} | X_{n-2}, \dots, X_1) \cdots P(X_2 | X_1) \cdot P(X_1) = \\ \prod_i P \left[ X_i | \prod_{X_i} \right] \end{aligned} \quad (4)$$

### 3.3 贝叶斯网络的构建方法

对于不同的应用领域构造贝叶斯网络的方法不尽相同,但概括起来包括以下三个步骤:

(1)确定特征变量。对所要为之建模的问题领域,确定所有的特征变量,这些变量构成该问题领域的变量集,对变量集中的每个变量确定其状态和取值范围<sup>[3]</sup>。

(2)确定网络结构。根据节点之间的概率依赖关系或先验依赖关系确定网络结构,对具备大量专家知识的问题领域,贝叶斯网络结构的构建可以由先验知识获得。大多数情况下由专家知识获得的贝叶斯网络结构正是最优的贝叶斯网络结构。

在领域专家知识或先验知识难以获得的情况下,一种更普遍的方法是直接从大量的样本数据中主动地学习出节点之间的关系构成贝叶斯网络结构,这种方法称为从数据中学习贝叶斯网络结构。

(3)计算每个节点的概率分布。对确定的网络结构计算每个节点的概率分布,包括根节点的先验概率分布和中间节点的条件概率分布。贝叶斯网络结构确定以后,贝叶斯网络的参数学习就变为在给出网络结构和样本数据集的条件下计算节点的概率分布问题。

## 4 软件项目风险评估

软件项目在系统分析和设计环节中经常存在项目计划制定不完善、需求不确定、技术或方法选择不恰当等因素,并且这些因素相互影响大,对风险的评估尤为重要。以制造业信息化公共服务平台项目为例,说明如何在软件项目中利用贝叶斯网络进行风险评估。

### 4.1 贝叶斯网络的建立

制造业信息化公共服务平台属于创新型项目,项目组在研发过程中需求变更比较频繁,开发人员对开发平台与开发技术难以把握,对平台资源整合以及业务流程自动化的可视化工作流引擎实现难度大,对软件性能的提高有较大的影响。

在项目启动时实施风险评估,首先识别项目中的风险和风险因子,设定风险关键字,即需求风险、需求变更、技术风险、技术熟悉度、软件复杂性、人员沟通、软件性能以及安全策略等。根据已有案例启动风险案例学习机制,结合专家经验生成本次评估的贝叶斯网络即风险分析网络,如图2所示。

为与 $B_m$ 的定义相契合,在图2所示中用节点变量 $X_i$ 代替风险分析网络中的节点,得到如图3所示的简化贝叶斯网络结构。

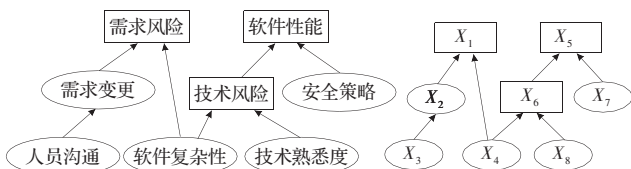


图2 贝叶斯网络结构图 图3 简化的贝叶斯网络结构图

根据贝叶斯网络的形式化分析,则有:

$$N_r = \{X_1, X_5, X_6\} \quad (5)$$

$$N_f = \{X_2, X_3, X_4, X_6, X_7, X_8\} \quad (6)$$

$$R = \{ \langle X_2, X_1 \rangle, \langle X_3, X_2 \rangle, \langle X_4, X_1 \rangle, \langle X_4, X_6 \rangle, \langle X_8, X_6 \rangle, \langle X_6, X_5 \rangle, \langle X_7, X_5 \rangle \} \quad (7)$$

$$P(X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8) = P(X_1 | X_2, X_4) \cdot P(X_2 | X_3) \cdot P(X_3) \cdot P(X_4) \cdot P(X_5 | X_6, X_7) \cdot P(X_6 | X_4, X_8) \cdot P(X_7) \cdot P(X_8) \quad (8)$$

### 4.2 风险概率的计算

软件项目风险评估是软件项目风险管理过程中的核心,评估过程分为风险发生概率评估和风险影响评估。

#### 4.2.1 风险发生概率评估

风险分析过程是依据贝叶斯网络进行的,贝叶斯网络是表示变量间概率依赖关系的有向无环图,每个节点表示领域变量,每条边表示变量间的概率依赖关系,同时每个节点都对应着一个CPT,指明该变量与父节点之间概率依赖的数量关系。风险发生概率的评估步骤为:

(1)定义风险发生概率的模糊评语集。在缺乏项目历史数据的情况下,领域专家将使用“不可能”、“可能”等模糊性语言来评价风险发生的概率,即模糊评语集 $H_p = \{\text{极不可能, 不可能, 中等, 可能, 很可能}\}$ <sup>[4]</sup>。

(2)风险发生概率评估是指专家使用模糊评语集 $H_p$ ,设定风险分析网络中的条件概率和先验概率,形成CPT。图2中软件性能风险节点的CPT设置如表1所示。

表1 软件性能风险节点的CPT设置

技术风险	安全策略	软件性能=true
0	0	0.02
0	1	0.30
1	0	0.45
1	1	0.90

(3)根据风险分析网络和CPT,通过风险概率推理算法,计算得出各个风险节点的发生概率综合评估矩阵 $P$ 。

#### 4.2.2 风险影响评估

风险影响评估包括两个部分,即对风险损失的评估和风险影响的综合评估。风险影响的评估步骤为:

(1)定义风险因素后果集 $D = \{\text{进度, 成本, 软件质量}\}$ ,风险对后果影响的模糊评语集 $H_c = \{\text{极低, 低, 中等, 高, 极高}\}$ 。

(2)风险损失评估。专家使用模糊评语集 $H_c$ 构建模糊专家评估矩阵 $C$ ,以此来评价各种风险对后果集造成的影响。 $C$ 由 $M$ 个风险后果, $N$ 类风险组成,任意一个 $C_{ij}$ 表示第 $i$ 类风险对第 $j$ 类风险后果的影响做出的评估。制造业信息化服务平台风险损失评估如表2所示。

表2 风险损失评估

风险/损失/后果	进度	费用	软件质量
需求风险	高	中等	极低
技术风险	高	中等	极高
软件性能	中等	中等	极高

根据风险评估损失以及风险节点发生概率,调用风险当量的计算公式 $R = P * C$ ,得到风险当量的精确数据如表3所示,对应的图示表示如图4所示。

表3 风险当量量化数据

风险/损失/后果	进度	费用	软件质量
需求风险	0.47	0.24	0.03
技术风险	0.54	0.36	0.78
软件性能	0.25	0.25	0.67

(3)风险权重计算。在不同的项目、不同的环境及不同的资源下,风险因素的重要程度有所不同,用风险因素的权重系数向量 $A = [a_1, a_2, \dots, a_n]^T$ 表示。其中 $n$ 为风险因素个数, $a_i$ 为第 $i$

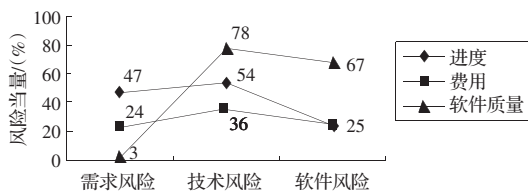


图4 风险当量图示

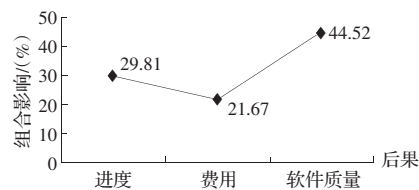


图6 风险组合影响

类风险因素的相对重要程度。 $a_i$ 可由AHP法求出,具体方法是:①将不同的风险因素列成比较矩阵;②按照表4中的1~9标度法进行两两比较;③由方根法求其权重值,归一化处理得到向量A,并求出一致性指标CR;④当 $CR < 0.1$ 时,该比较矩阵满足一致性要求,否则需调整比较矩阵的值<sup>[9]</sup>。风险权重计算结果如表5所示。

表4 1~9 标度法及含义

取值	比较结果的量化
1	同样重要
3	稍重要
5	明显重要
7	重要得多
9	极端重要
2,4,6,8	上述两相邻判断的折中
上述各数的倒数	反比较

表5 风险权重评估量化与计算结果

风险/权重评估	需求风险	技术风险	软件性能
需求风险	1.00	0.33	0.14
技术风险	3.00	1.00	0.17
软件性能	6.99	5.59	1.00

(4)风险影响综合评估。风险当量是评估软件风险程度的关键指标,目前的分析方法通常由风险概率与后果的乘积度量出单个风险对各风险后果的危害程度,而多种风险共同造成的损失以及风险对整体后果的影响却无法体现。为了从多维的视角更深入地了解风险,以更好地达到控制风险的目的,采用如下方法来解决这一问题:

$$R_{\text{组合}} = [R_1, R_2, \dots, R_m]^T, \text{其中 } R_i = \sum_{j=1}^n a_i p_j c_{ij} \quad (9)$$

$$R_{\text{综合}} = [R_1, R_2, \dots, R_n] / (R_1 + R_2 + \dots + R_n), \text{其中 } R_i = p_j \sum_{i=1}^3 c_{ij} \quad (10)$$

根据风险损失评估结果和风险权重,计算出风险综合影响和组合影响值,风险综合影响如图5所示,风险组合影响图6所示。

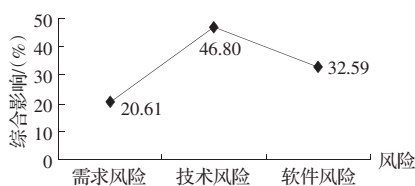


图5 风险综合影响

## 5 结束语

实践证明,在软件项目中引入贝叶斯网络进行风险评估能够有效地预测潜在的风险。软件项目风险发生的概率通过贝叶斯网络进行推理,降低了评估的难度和主观性;使用模糊性语言评估风险后果及损失,解决了专家评估的不确定性问题;根据评估结果,可以有效地提供风险预防控制措施,能够避免和减少风险造成的损失;同时,贝叶斯网络具有完善的学习机制,能够逐步丰富软件企业的风险数据库,在不断的学习和修正过程中提高了风险的预测和应变能力,为有效地降低风险发生概率、提高软件开发成功率提供了一种新的途径,具有很好的应用价值。

软件项目风险评估体系的建立是以企业长期持续的软件风险评估为基础的,需要一个日趋完善的风险案例库,案例库是随着评估案例的增多而逐渐丰富起来的。因此,如何快速建立初始风险案例库还要依赖于领域专家知识的支持。风险评估系统中风险因果关系的分析是以无反馈循环为前提的,而现实世界中的因果关系往往是循环的,在将来的工作中拟引入时间序列来模拟这种反馈循环,尽可能地接近现实。

## 参考文献:

- 唐爱国,王如龙.基于贝叶斯网络的软件项目风险评估模型[J].计算机工程,2008,34(22):91-93.
- 潘春光,陈英武,汪浩.软件项目风险管理理论与方法研究综述[J].控制与决策,2007,22(5):481-486.
- Bannerman P L.Risk and risk management in software projects: A reassessment[J].Journal of Systems and Software,2008,81(12):2118-2133.
- 李美华,付宏.软件项目风险评估模型的建立[J].吉林大学学报,2005,23(6):696-701.
- 蒋国萍,陈英武.基于面向对象贝叶斯网络的软件项目风险评估[J].系统工程与电子技术,2005,27(2):353-356.
- 范敏,黄席隼,石为人,等.一种改进的贝叶斯网络结构学习算法[J].系统仿真学报,2008,20(17):4613-4617.
- 袁晓峰,陈淑红,许化龙.贝叶斯网络推理的一种高精度仿真算法[J].系统仿真学报,2009,21(1):108-111.
- Langseth H,Portinale L.Bayesian networks in reliability[J].Reliability Engineering and System Safety,2007,92(1):92-108.
- 张永铮,方滨兴,迟悦.用于评估网络信息系统的风险传播模型[J].软件学报,2007,18(1):137-145.
- Fenton NE,Radliński Ł,Neil M.Improved bayesian networks for software project risk assessment using dynamic discretisation[C]//IFIP Working Conference on Software Engineering Techniques: Design for Quality,2006,227:139-148.