

一种基于贝叶斯网络的软件项目风险管理方法^①

李天纵, 王 强

(北京航空航天大学 经济管理学院, 北京 100083)

摘 要: 提出一种新的软件项目风险管理方法, 采用贝叶斯网络同时对风险发生概率和风险影响进行推理。该风险管理方法能够随着软件项目的进行持续地评估潜在风险, 并采取适当的措施应对风险。实践证明, 在软件开发过程中引入该风险管理方法能够有效地对风险进行管理, 提高软件开发的成功率。

关键词: 贝叶斯网络; 软件项目; 风险分析; 风险评估; 风险管理

A Method of Software Project Risk Management Based on Bayesian Networks

LI Tian-Zong, WANG Qiang

(School of Economics and Management, Beihang University, Beijing 100083, China)

Abstract: A new method of software project risk management is presented, which uses the Bayesian networks to infer the probability of risk occurrence and the impact of risk simultaneously. The method can assess risks continuously and take proper measures to respond to risks during software project. Practice proves that by introducing this method into software development process, project risks will be well-managed and project success rate will be improved.

Keywords: Bayesian networks; software project; risk analysis; risk assessment; risk management

1 引言

项目的创新性、一次性、独特性及其复杂性决定了项目风险的不可避免性; 风险发生后损失的难以弥补性和工作的被动性决定了风险管理的重要性。随着技术更新和产业环境变化, 软件项目趋向于周期长、规模大、涉及范围广等特点, 使得项目涉及的风险数量众多, 各种风险之间的内在关系错综复杂。因此, 科学地分析风险间的关系, 并基于此对软件项目进行有效的风险管理很有必要。

软件工程协会提出的软件项目风险管理经典模式——连续风险管理, 将风险管理划分为 5 个过程, 并且组成循环往复的圆周。这五个过程是: 风险识别、风险分析、风险计划、风险跟踪和风险应对^[1]。

以此为基础本文提出了一种基于贝叶斯网络的软件项目风险管理模型, 该模型不仅全面涵盖了这 5 个基本过程元素, 而且其整个风险管理活动是在一个具有持续反馈功能的风险管理流程当中进行的, 能够实现软件项目的动态风险管理。

2 贝叶斯网络技术基础

贝叶斯网络, 又称贝叶斯信念网络(Bayesian Belief Networks), 它是基于概率推理的图形化网络, 是贝叶斯方法的拓展。

一个贝叶斯网络可表示为:

$$B = \langle S, P \rangle \quad (1)$$

它包含一组随机变量 $X = \{X_1, X_2, \dots, X_n\}$ 。其中, S 表示网络的拓扑结构, 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随机变量, 每个节点可以有若干状态, 代表随机变量可取的值, 图中的有向弧表示随机变量间的因果依赖关系; P 表示贝叶斯网络中节点间的条件概率或先验概率集合, 即节点概率表(NPT)。

对于网络中任一节点 X_i , 其双亲节点定义为, 指向该节点的弧的出发节点。将 X_i 的双亲节点集合表示为 $Pa(X_i)$, 则 X_i 的条件概率为 $P(X_i | Pa(X_i))$, 指明了 X_i 在的双亲节点的状态确定时, X_i 处于某一状态的可能性。另外, 贝叶斯网络还给出了随机变量集 X 上的联

^① 收稿时间:2010-05-17;收到修改稿时间:2010-06-21

合概率分布:

$$P(X_1, X_2, \dots, X_n) = \prod_{i=1}^n P(X_i | Pa(X_i)) \quad (2)$$

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

3 软件项目风险管理模型

与多数基于贝叶斯网络的软件项目风险管理模型相比, 本文提出的模型对于风险影响的评估也是利用贝叶斯网络推理而得, 其合理性在于: 存在联系的风

险, 它们各自发生时对项目的影响也必然存在相关性。以往基于贝叶斯网络的风险管理模型只能对风险发生概率进行智能化持续评估, 对于各风险的影响值的评估, 有的是一次性的, 不能随着项目的进行而调整, 有的在每次调整时都需要再次收集专家知识, 这样会带来极大的工作量, 需要很长时间才能确定新的风险影响值, 甚至会跟不上项目风险状态的改变。本模型最突出的优势在于: 对风险发生概率与风险影响共同进行全程智能化持续评估。

本文提出的基于贝叶斯网络的软件项目风险管理模型如图 1。

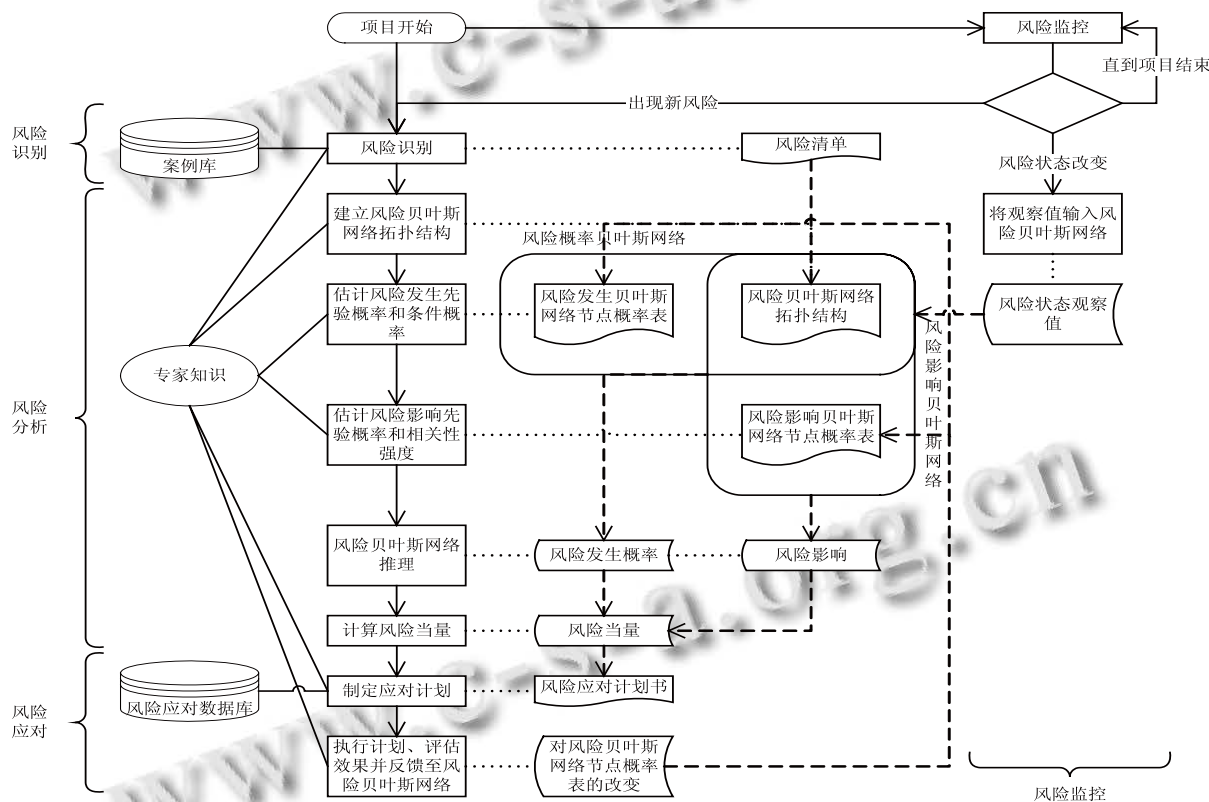


图 1 基于贝叶斯网络的软件项目风险管理模型

3.1 风险识别

由于本文研究的是软件项目的风险, 因此需要对于软件开发过程各阶段特有的项目风险(时间维风险), 以及独立于时间的、存在于多个项目阶段的各种风险(空间维风险)加以识别, 这依赖于项目管理者或专家的知识, 同时可以结合案例库进行。风险识别的结果是软件项目风险清单。

风险识别不仅在项目开始前进行, 项目全程都需要监视新的风险的诞生, 使得模型能将项目全程任何时刻的重要不确定性都纳入管理范围, 进行系统化管理。

3.2 风险分析

本模型是用风险贝叶斯网络对风险列表中的各项风险进行评估, 从而指导项目团队制定优先决策, 为项目管理者专注于重要的风险提供行动依据。因此首

先需要构建风险贝叶斯网络——分为进行风险发生概率推理的风险概率贝叶斯网络,以及进行风险影响推理的风险影响贝叶斯网络。由于两个网络的任一节点都对应于一个项目风险,而任一条弧对应于一个风险因果关系,因此两个网络的网络拓扑结构是相同的,不同在于:风险概率贝叶斯网络的每个节点存在两个状态——“发生”与“不发生”,而风险影响贝叶斯网络的每个节点则存在多个状态,分别对应不同的影响等级;并且相应的节点先验概率和条件概率信息的定义也不相同。

3.2.1 建立风险贝叶斯网络拓扑结构

建立风险贝叶斯网络的拓扑结构,实际上就是发掘已识别的各个风险间存在的因果关系。本模型提供两种风险贝叶斯网络建立方法。

基于统计分析和专家知识的拓扑结构获取法:将风险清单中的风险划分为通用风险和项目特有风险,通用风险指普遍存在于软件开发项目中的风险,特有风险指所管理的特定软件项目特有的风险。对于通用风险进行问卷调查,被调查人需要选择其认为常见的风险,之后对问卷调查结果进行统计分析,分析风险相关性,相关性较强的风险间通过已有知识确认其是否确实存在因果联系,从而得到通用风险贝叶斯网络拓扑结构,它是各个软件项目通用的基础结构。此外需要利用专家知识将特有风险加入通用结构从而最终得到该项目的风险贝叶斯网络拓扑结构。

基于案例库的拓扑结构获取法^[2]:如果有相似案例的历史数据作为基础,风险评估工作相对会容易很多,因为可利用历史案例协助建立风险分析模型,达到模型重用的目的。具体做法是:在案例库中搜索包含当前项目风险节点的案例;计算当前项目风险节点和各案例的相似度和背离度;筛选出相似度最大、背离度最小的案例;将筛选出的案例进行剪枝和重组,形成满足当前项目要求的贝叶斯网络结构,存为风险贝叶斯网络结构文件。需要指出的是,在项目结束后,还要将最终的风险贝叶斯网络结构等信息存入案例库,这样即可逐渐形成一个包含诸多独立风险分析模型的案例库,为以后项目风险贝叶斯网络的建立提供参考。

3.2.2 构建风险概率贝叶斯网络

由贝叶斯网络的定义可知,构建一个完整的贝叶斯网络,除了建立其网络拓扑结构之外,还需要获取其上

的概率信息,包括网络边缘节点的先验概率信息以及非边缘节点的条件概率信息,它们将以风险概率贝叶斯网络节点概率表(RPBN-NPT)的形式存储,与风险贝叶斯网络结构文件一起完整描述风险概率贝叶斯网络。

风险概率贝叶斯网络中每个风险节点有两个状态——“发生”与“不发生”,需要获取的概率信息是指每个节点分别处于这两个状态的先验概率或条件概率信息。信息的获取依赖于专家知识,假设有 n 位专家, w_i 代表第 i 位专家的归一化权重,满足。对于边缘风险节点 R_j ,第 i 位专家认为其发生的概率为 $P_i(R_j)$,则该节点发生的先验概率定义为:

$$P(R_j) = \sum_{i=1}^n w_i P_i(R_j) \quad (3)$$

对于非边缘节点 R_k ,将其所有双亲变量用集合 $Pa(R_k)$ 的形式表示。已知其双亲风险节点发生与否的信息时,第 i 位专家认为其发生的条件概率为 $P_i(R_k|Pa(R_k))$,则该节点发生的条件概率定义为:

$$P(R_k|Pa(R_k)) = \sum_{i=1}^n w_i P_i(R_k|Pa(R_k)) \quad (4)$$

3.2.3 构建风险影响贝叶斯网络

风险在未发生前,其一旦发生给项目带来的影响实际上也是不确定的,因此在估计风险影响时应该将这种不确定性考虑进来。相互联系的风险,它们发生时对于项目目标的影响也应该是相关的,当某个风险确实发生从而确定其实际造成的影响时,以及当执行一个风险应对计划使风险间联系程度发生改变时,对关联风险影响程度的估计都可能发生变化——构建风险影响贝叶斯网络就是为解决这样的问题。

这里只需获得风险影响贝叶斯网络节点概率表(RIBN-NPT)。相对复杂在于,风险影响贝叶斯网络的各个节点有多个状态,若使用构建风险概率贝叶斯网络的方法,则在估计每个风险的影响时,需要估计的条件概率数目会非常多。为了在保证合理性的前提下最大程度地减少工作量,使模型实用性更强,需要设计专门的构建方法。

对于边缘风险节点 R_j ,其风险影响为 I_j ,第 i 位专家认为其发生后对项目目标的影响程度为 D_{ij} , $D_{ij} \in \{VL, L, M, H, VH\}$,则该风险节点影响程度处于各等级的先验概率定义为:

$$P(I_j = VL) = \sum_{D_{ij}=VL} w_i$$

$$\vdots$$

$$(5)$$

$$P(I_j = VH) = \sum_{D_{ij}=VH} w_i$$

满足 $\sum_{c \in \{VL, L, M, H, VH\}} P(I_j = c) = 1$ 。如此一来, 每位专家

对于每个边缘风险节点的影响只需给出一个确定性的判断即可, 降低了对专家的要求, 获得的先验概率信息更可靠, 关键是提高了收集和处理专家知识的效率。

尽管如此, 这步工作大部分复杂性仍旧体现在大量的非边缘节点的条件概率信息获取之上。本模型引入了一种利用双亲节点对子节点影响程度生成条件概率的方法, 这种方法只需要知道各个双亲节点与子节点相关性的相对强弱即可。假设风险节点 R_a 有两个双亲节点 R_b 、 R_c , R_a 与二者的归一化相关性强度分别为 w_{ab} 、 w_{ac} , 满足 $w_{ab} + w_{ac} = 1$, 那么利用 w_{ab} 和 w_{ac} 就能够生成一系列所需的条件概率。

这一方法所需的归一化相关性强度的获得, 借鉴了层次分析法中 1~9 标度下的两两比较法。将第 i 个专家对上述归一化相关性强度的判断记为: $w_{ia} = [w_{iab}, w_{iac}]$; 将所有专家的判断整理为对 R_a 的专家判断矩阵: $J_a = [w_{1a}, w_{2a}, \dots, w_{na}]T$; 将专家权重整理为向量形式: $W = [w_1, w_2, \dots, w_n]$ 。则综合所有专家知识的风险节点 R_a 与其双亲节点 R_b 、 R_c 的归一化相关性强度为:

$$W_a = [w_{ab}, w_{ac}] = W \cdot J_a \quad (6)$$

引入这一方法充分降低了对于专家的要求, 极大地减少了获取非边缘风险节点影响程度的条件概率的工作量, 也真正使得利用贝叶斯网络进行风险影响推理变得可行。

3.2.4 风险贝叶斯网络推理

利用完整构建的风险概率贝叶斯网络和风险影响贝叶斯网络的推理能力, 即可对各个风险进行智能化持续评估: 在项目进行中, 只要监测到有风险发生, 就把该风险发生的信息以及其所造成的实际影响程度信息分别输入两个网络中, 网络会进行信念更新, 从而协助项目团队制定最佳决策进行风险应对。对于风险节点 R_j , 其期望影响为 $E(I_j)$, 由风险概率贝叶斯网络可知其发生概率为 $P(R_j)$, 则该风险的风险当量 D_j 为:

$$D_j = P(R_j) \cdot E(I_j) \quad (7)$$

3.3 风险应对

制定风险应对计划: 既可以直接调用风险应对数据库, 也可以依赖专家知识制定新计划, 或是将二者结合进行。应用风险应对数据库的方法更加智能化和便捷, 但很多时候, 由于风险间错综复杂的关系, 完全按照风险当量大小的顺序处置风险未必是最佳策略, 往往需要人的介入, 才能更有效率。因此, 有风险应对数据库支持的、结合专家知识的应对计划制定方法是值得推崇的。

执行风险应对计划: 执行力的强弱直接关乎项目风险管理的效果。风险应对的目的就是将当下较严重的风险降低到可以承受的范围内, 所以需要反馈信息给风险贝叶斯网络, 以体现应对效果。这里所需的反馈信息并不是自动产生的, 它与应对计划的执行效果有关, 可以看作专家对执行效果的评价, 表现为一系列概率值和相关性强度的改变。

3.4 风险监控

项目管理者需要随时关注项目各种内外部的变化, 及时发现新的风险并加以识别, 进而保证模型时刻将全部相关风险纳入管理范围; 对于已有风险, 需要密切留意其状态, 一旦确认其发生, 须立刻对其实际造成的后果进行评价, “通知” 风险概率贝叶斯网络它确已发生并 “告知” 风险影响贝叶斯网络它的实际影响, 使得风险贝叶斯网络能够发挥其信念更新的能力, 指导项目团队快速响应项目风险, 从而保证开发工作向着目标稳步进行。正是由于风险监控的存在, 项目风险管理才得以成为一个循环往复的过程。

4 结束语

实践证明, 在软件开发过程中引入本模型能够较为有效地预测风险发生的概率以及可能造成的影响, 从而指导项目团队随时对工作进行调整, 以保证项目向既定目标稳步进行。另外, 在运用该风险管理模型的过程中, 贝叶斯网络中各风险节点间条件概率的确定以及风险的量化方法还有待进一步研究; 同时, 风险影响贝叶斯网络建模中引入的利用相关性强度代替条件概率的方法, 也有待进一步检验。

参考文献

- 1 Feng N, Li MQ, Kou JS, Fang DY. Software project risk management model based on Bayesian networks. *Computer Engineering*, 2007, 33(7): 41-43.
- 2 Tang AG, Wang RL. Software project risk assessment model based on Bayesian networks. *Computer Engineering*, 2008, 34(22): 91-93.