

基于信息熵的软件开发风险度量模型^{*}

姜 茸^{1,2}, 廖鸿志², 张新明¹, 曾志勇¹

(1. 云南财经大学 信息学院, 云南 昆明 650221;

2. 云南大学 软件学院, 云南 昆明 650091)

摘 要: 给出一个定量的基于信息熵的软件开发风险度量模型, 该模型考虑了由于风险因素带来的损失及其发生概率, 实践表明该模型能有效地综合度量软件开发风险。

关键词: 风险度量; 信息熵; 软件开发; 软件工程

中图分类号: TP311.5

文献标识码: A

Model of measuring software development risk based on information entropy

JIANG Rong^{1,2}, LIAO Hong Zhi², ZHANG Xin Ming¹, ZENG Zhi Yong¹

(1. School of Information, Yunnan University of Finance and Economics, Kunming 650221, China;

2. School of Software, Yunnan University, Kunming 650091, China)

Abstract: This paper provides a quantificational model to measuring the software development risk based on information entropy. The model involves both the probabilities of the risk factors and the loss, it is an effective synthesis method to measuring software development in practice.

Key words: risk measure; information entropy; software development; software engineering

风险就是客观存在的不确定性以及可能给项目带来的损失, 风险对软件项目成败与质量有很大的影响, 较早识别并准确度量软件开发风险可以有效地降低或克服风险的危害, 度量软件开发风险无论是对软件企业还是软件工程学科都有十分重要的意义和作用。但是, 软件项目不同于一般工程项目, 软件系统是一种概念系统, 它是一种逻辑产品, 因此软件开发风险度量比其他工程项目风险度量更加困难。

目前, 国内外风险评估和度量的方法很多, 其分析侧重点、使用范围各有不同, 但定量的度量方法不多。常用的方法有基于模糊数学、神经网络、灰色理论、层次分析、构造风险矩阵等方法。“模糊综合评价法”对各风险因素权值的确定一般采用专家主观评定, 权值误差的积累会导致风险分析结果偏离实际而且不利于综合定量分析; 基于神经网络或灰色理论的评估方法计算显得过于复杂; 层次分析

方法又不太适合有较多平行风险因素的度量; 参考文献[1]基于信息熵的信息安全风险模型, 它建立在故障树的基础上, 引入了信息熵, 但此方法是分析系统的信息安全而非软件开发过程的风险度量, 而且此方法仅适用于层次结构模型。本文提出一种定量的基于信息熵的软件开发风险度量方法, 该方法不仅实用而且计算简单。

1 信息熵作为风险度量方法的理论根据

信息熵是 Shannon 1948 年于“通信的数学理论”中提出的概念, 认为信息即通信前后的不确定性之差, 定义的 n 个可能的消息 X_1, X_2, \dots, X_n , 其发生的概率分别为 P_1, P_2, \dots, P_n , 且 $0 < P_i \leq 1 (i=1, 2, \dots, n)$, $\sum_{i=1}^n P_i = 1$, 函数 $H(P_1, P_2, \dots, P_n) = -k \sum_{i=1}^n P_i \ln P_i$ (其中 k 为比例系数), 用来表示信源输出后每个消息提供的平均信息量, 或信源输

* 基金项目: 云南省科技计划项目([2008]016); 云南省教育厅科学研究基金项目([2009]06)

技术与方法 Technique and Method

出前的平均不确定程度,他将此函数称为信息熵。既然把信息定义为通信中消除了的不确定性,信息的度量就是对这种不确定性的度量^[2],信息熵能有效地度量通信中消息的信息量,熵值越大,不确定性越大,信息量越小。

虽然对风险的含义有各种理解,但风险表现出的是一种不确定性,这是风险的本质,不确定性越大,风险也越大,与信息熵不谋而合。因此可以利用信息熵度量软件开发的^[3]风险,不确定性越大,熵值越大,风险就越大。

2 基于信息熵的风险度量模型

2.1 风险定义

软件开发风险包含不确定性和损失 2 个重要特性。风险可定义为 1 个三元组 $R=(X, P, C)$,其中 X 表示风险因素集合, P 表示风险发生的概率, C 表示风险产生的损失。进行风险度量时关键是量化不确定性的程度和每个风险相关因素发生时的损失程度^[3]。

2.2 信息熵量化风险

不确定性是风险的重要特性,但风险 \neq 不确定性, $C=$ 它还涉及风险因素发生时产生的损失程度。发生可能性越大的风险因素,给项目带来的风险越高;发生时损失越大的风险因素,给项目带来的风险越大。因此度量风险大小既要^[4]看风险因素发生的概率也要看它对软件项目产生的影响,可将风险大小(或称为影响程度)定义为发生概率与损失程度之积。

为此,设软件开发项目存在 n 个风险因素,分别为 X_1, X_2, \dots, X_n ;其发生概率为 $P_1, P_2, \dots, P_n; 0 < P_i \leq 1$,损失度为 $C_1, C_2, \dots, C_n; 0 < C_i \leq 1$,于是,风险影响度 A_1, A_2, \dots, A_n :

$$A_i = \sqrt{P_i C_i}, i=1, \dots, n \quad (1)$$

式中, $0 < A_n \leq 1$ 。将 A_i 归一化处理,可得:

$$\rho_i = \frac{A_i}{\sum_{j=1}^n A_j}, i=1, \dots, n \quad (2)$$

式中, $0 < \rho_i \leq 1, \sum_{i=1}^n \rho_i = 1$

由风险定义及信息熵理论,得出如下基于信息熵的风险度量模型:

$$R = - \sum_{i=1}^n A_i \frac{\rho_i \log_2 \rho_i}{\log_2 n} \quad (3)$$

式中, $0 < R \leq 1$ 。

3 模型实例分析

这里对 7 个不同软件项目进行风险度量,参考文献 [4]总结了 16 种风险因素,其中 10 个为主要风险因素:需求与变更风险、预算风险、技术风险、支持工具风险、维护风险、软件过程能力风险、进度风险、资源风险、市场风险、营销风险,各个因素发生的概率及其损失度如下:

$P=$	0.1	0.2	0.5	0.3	0.7	0.6	0.9	0.55	0.45	0.4
	0.8	0.6	0.4	0.8	0.5	0.45	0.2	0.6	0.5	0.4
	0.75	0.3	0.5	0.8	0.8	0.85	0.9	0.95	0.8	0.55
	0.8	0.9	0.75	0.85	0.8	0.95	0.88	0.75	0.9	0.85
	0.7	0.75	0.2	0.9	0.9	0.8	0.8	0.4	0.5	0.55
	0.35	0.7	0.1	0.3	0.65	0.3	0.5	0.45	0.4	0.75
	0.5	0.4	0.3	0.6	0.55	0.35	0.65	0.7	0.1	0.6
	0.8	0.75	0.3	0.6	0.8	0.75	0.2	0.65	0.4	0.3
	0.5	0.2	0.5	0.7	0.55	0.6	0.2	0.3	0.25	0.45
	0.5	0.4	0.65	0.2	0.1	0.3	0.1	0.15	0.3	0.7
	0.9	0.85	0.8	0.95	0.75	0.8	0.85	0.7	0.8	0.9
	0.5	0.4	0.3	0.1	0.2	0.3	0.45	0.7	0.75	0.6
	0.5	0.4	0.8	0.75	0.6	0.45	0.35	0.2	0.4	0.55
	0.25	0.3	0.6	0.2	0.25	0.2	0.6	0.35	0.45	0.3

其中,矩阵 P 中第 i 行的 10 个元素为第 i 个项目的 10 个风险因素发生的概率,矩阵 C 中第 i 行的 10 个元素为第 i 个项目的 10 个风险因素发生时的损失程度。

根据(1)式,可得:

$A=$	0.282 8	0.387 3	0.387 3	0.424 3	0.748 3	0.670 8	0.424 3	0.597 9	0.424 3	0.346 4
	0.632 5	0.346 4	0.447 2	0.748 3	0.524 4	0.519 6	0.2	0.424 3	0.353 6	0.424 3
	0.612 4	0.346 4	0.570 1	0.4	0.282 8	0.505	0.3	0.377 5	0.489 9	0.620 5
	0.848 5	0.874 6	0.774 6	0.898 6	0.774 6	0.871 8	0.864 9	0.724 6	0.848 5	0.874 6
	0.591 6	0.547 7	0.244 9	0.3	0.424 3	0.489 9	0.6	0.529 2	0.612 4	0.574 5
	0.443 3	0.529 2	0.282 8	0.474 3	0.624 5	0.367 4	0.418 3	0.3	0.4	0.642 3
	0.353 6	0.346 4	0.424 3	0.346 4	0.370 8	0.364 6	0.624 5	0.495	0.212 1	0.424 3

由(2)式,可得:

$\rho=$	0.060 3	0.082 5	0.082 5	0.090 4	0.159 4	0.142 9	0.090 4	0.127 4	0.090 4	0.073 8
	0.136 9	0.075	0.096 8	0.162	0.113 5	0.112 5	0.043 3	0.091 8	0.076 5	0.091 8
	0.135 9	0.076 9	0.126 6	0.088 8	0.062 8	0.112 1	0.066 6	0.083 8	0.108 8	0.137 7
	0.101 6	0.104 7	0.092 7	0.107 5	0.092 7	0.104 3	0.103 5	0.086 7	0.101 6	0.104 7
	0.120 4	0.111 5	0.049 8	0.061	0.086 3	0.099 7	0.122 1	0.107 7	0.124 6	0.116 9
	0.093 9	0.118 7	0.063 5	0.106 4	0.140 1	0.082 4	0.093 9	0.067 3	0.089 7	0.144 1
	0.091 5	0.089 7	0.109 9	0.089 7	0.096	0.068 5	0.161 7	0.128 2	0.054 9	0.109 9

由风险度量模型公式,可求得这 7 个软件项目的开发风险分别为 $R=(0.483 4, 0.478 2, 0.461 6, 0.836 8, 0.503 3, 0.456 1, 0.396 6)$,其中第 4 个项目开发风险最大,第 7 个最小。

于是,软件开发商可以根据此风险度量模型结果采取一系列措施应对开发中的各种风险,如果该项目风险

过大且利润不高,甚至可以放弃项目的开发。

信息熵反映了系统的不确定性,可作为软件开发风险的度量方法;软件开发风险由不确定因素发生概率及这些因素发生时对项目产生的损失两方面决定。本文提出了一个基于信息熵理论的定量风险度量模型,该模型既考虑风险因素发生概率又考虑其损失,实践表明,该模型能有效地综合度量软件开发的

参考文献

- [1] 汤永利,徐国爱,钮心忻,等.基于信息熵的信息安全风险分析模型[J].北京邮电大学学报,2008(4):50-53.
- [2] 许国志,顾基发,车宏安.系统科学[M].上海:上海科技教育出版社,2000.
- [3] PRESSMAN R S.Software engineering-a practitioner's approach(fifth edition)[M].北京:清华大学出版社,2001.
- [4] 潘梅森,颜君颜.基于LVQ的软件项目风险评估模型的研究[J].计算机工程与应用,2007(12):126-130.

(收稿日期:2009-04-01)

