文章编号：2096-1472(2016)-01-32-03

一种可信属性之间的相关性分析方法

李俊霖[1]，高 涛[2]，郁 湧[3]
（1.北京市经济信息中心，北京 100031；
2.云南大学数统学院，云南 昆明 650091；
3.云南大学软件学院，云南 昆明 650091）

摘 要：随着软件系统在信息社会中发挥越来越重要的作用，人们对软件系统的可信性方面的要求也愈来愈高。软件可信性可由软件可信属性集来进行描述和表示，但是人们根据软件系统的不同特征和不同层面提出了很多软件可信属性集。为此，本文提出一种软件可信属性的表示以及相关性分析的方法，给出了可信属性相关性、独立性和冗余性的概念，并基于相关性分析可以从所知的属性集中构建一个适合的可信属性最小子集作为对软件可信性进行评估的指标。此方法能够很好地对可信软件属性之间的相关性进行分析与度量。

关键词：软件可信性；可信属性；相关性分析
中图分类号：TP31 **文献标识码**：A

An Approach to Correlation Analysis of Trustworthy Attributes

LI Junlin[1],GAO Tao[2],YU Yong[3]
(1.*Beijing Economic Information Center*,*Beijing* 100031,*China*;
2.*School of Mathematics and Statistics*,*Yunnan University*,*Kunming* 650091,*China*;
3.*School of Software*,*Yunnan University*,*Kunming* 650091,*China*)

Abstract: As the software plays more and more important roles in today's information society,the software reliability are strongly required.Software trustworthiness can be described and expressed by the trustworthy attributes,and many sets of trustworthy attributes are proposed based on the different features and aspects of the software systems.An approach to representation and correlation analysis of trustworthy attributes is proposed.Based on the correlation analysis,a suitable minimum subset can be given from the trustworthy attributes.And the minimum subset provides an effective method for measuring the software trust worthiness.The approach can be used to analyze and measure the correlation of trustworthy attributes.

Keywords:software trust worthiness;trustworthy attribute;correlation analysis

1 引言(Introduction)

可信软件作为软件领域最具挑战性和价值的研究课题之一，引起了国内外学者的高度重视。早在1991年，Laprie从安全关键系统的研究出发提出软件可依赖性(Dependability)的概念[1]。2001年，Algirdas等提出软件的可依赖性是指人们相信软件系统所具有的、特定服务的能力[2,3]，强调了人们对软件的信任。1997年美国国家科学技术委员会在《高可信系统的研究挑战》中明确提出了高可信性(High Confidence)的概念[4]。我国学者陈火旺、王戟等认为高可信软件系统在提供服务时能满足一系列关键性质[5]，具体涉及的性质包括：可靠安全性、可靠性、生存性、容错性、实时性、保密安全性中的一个或者多个。

德国科学研究基金委员会在奥尔登堡大学成立了可信软件研究院(Trust Soft Graduate School)，该机构研究认为，软件系统的可信性是由正确性、安全性、服务质量(包括性能、可靠性、可用性)、保密性及私密性所决定的[6]。刘克等认为可信软件(Trustworthy Software，TS)是指软件系统的动态行为及其运行结果总是符合人们预期，并在受到干扰时仍能提供连续服务的软件[7]。他们认为可信性是在正确性、安全性、可靠性、时效性、完整性、可用性、可预测性、生存性、可控性等众多概念基础上发展起来的一个新概念，是软件系统诸多属性的综合反映。本文作者也提出过一种基于构件的可信软件框架及其表示方法[8]。

2 可信软件和可信属性(Trusted software and trustworthy attribute)

在软件可信性的研究中，如何描述和度量软件的可信性

基金项目：国家自然科学基金项目“基于构件的可信软件构造及其行为动态可信测评”（61462091），云南省科技厅面上项目“基于构件的可信软件构造及其相关问题研究”（2012FB119，2012FB118），云南省教育厅科研重点项目“基于构件的可信软件构造及其可信测评”（2013Z057）和云南省软件工程重点实验室开放基金面上项目“基于构件的可信软件度量及失效波及效应分析”（2012SE305）.

是当前研究的一项重要内容。一般来说，软件的行为及其产生的结果可以通过一组适当的属性来描述，即软件可信性可依据能够反映软件某种可信性的一组属性以及用户在这组属性上的预期来共同刻画，这贯穿于软件的整个生命周期。为此，人们提出了很多软件可信属性类型和分类方法。

软件可信属性是一组用来描述和评价软件系统可信的特性，通过这些属性可以更明确具体地描述软件可信的内涵。Avizienis等[9,10]提出了可信性的概念框架，该框架中可信属性包括：可用性(availability)、可靠性(reliability)、防危性(safety)、机密性(confidentiality)、可维护性(maintainability)等。同时指出，软件的可用性、完整性和机密性构成了软件的安全性。陈火旺院士提出了高可信(high confidence)性质的概念[5]，认为软件可信性质包括：可靠性(reliability)、防危性(safety)、安全性(security)、可生存性(survivability)、容错性(fault to lerance)、实时性(real time)。

王怀民等主张从“客观性”和“主观性”两个方面分别定义可信软件：“软件可信性”指软件客观具有的质量；“可信软件”指用户对软件客观质量的主观认同。他们认为的影响软件可信性的属性包括：安全性、可用性、可靠性、实时性、可维护性和可生存性[11]。

此外，其他的研究者对可信性所包含的可信属性也有不同的见解，其中，大部分的研究者认为可信性应包括可用性(Availability)、可靠性(Reliability)、可维护性(Maintainability)、防危性(Safety)、保密性(confidentiality)、完整性(integrity)和可生存性(survivability)等属性，而每个属性特性又可能包含了若干的子特性，这些可信属性特征共同构成了软件的可信属性模型。

软件可信性是通过软件可信属性集及其取值来进行描述的，可信属性集是一组软件可信属性的集合，它是与软件可信相关的一组指标体系。可以通过用户判断的软件所具有的关于可信属性的评价和取值来判断软件系统的可信性，从而对软件系统的可信性进行不断的改进。在软件系统的运行过程中，若可信属性满足要求，则意味着软件达到应有的可信程度或该软件能达到其预设的可信目标；否则，就可以认为软件系统在运行过程中不可信。

软件可信性可由软件可信属性集来进行描述和表示，但是人们根据软件系统的不同特征和用户所关心的不同层面提出了很多描述软件可信的属性集。这些属性集中的属性可能不相同，有些是冗余的，有些甚至是相互冲突的，因此，首先要解决的问题是如何选择属性来描述软件的可信性，并且需要考虑这些所选择的属性之间是否有冗余，分析它们之间具有什么样的相关性。

3　可信属性相关性分析(Correlation analysis of trustworthy attributes)

3.1　可信属性的表示

对于一个可信软件系统TS，设其可信属性集为A={a1，a2，…，an}，V是可信属性值的集合，$V=\bigcup_{a\in A}V^{(a)}$，$V^{(a_i)}=[\underline{x_i},\overline{x_i}]$是可信属性$a_i$的值域，即可信属性$a_i$的取值范围，$f:\mathrm{TS}\times A\to V$是一个可信属性函数，它指定某一可信软件系统TS中可信属性的取值范围。由于软件系统的特征和用户所关心的层面不同，不同的可信软件，其属性取值范围也各不相同。

由于可信软件系统的运行特征，软件可信属性的取值也是一个动态的过程。随着时间的推移或者软件的运行，可信属性取值会不断发生变化，因此，可以把每个可信属性a_i的取值看成一个时间序列$\{a_i(\mathrm{t}),t=1,2,...,T\}(\underline{x_i}\le a_i(\mathrm{t})\le\overline{x_i})$，$a_i(\mathrm{t})$逐一记录了可信属性$t=1,2,...,T$在$a_i$时刻的取值。

3.2　可信属性之间的相关性分析

对于一个可信软件系统TS的可信属性集A中的任意两个可信属性a_i和a_j，如果它们在可信软件系统TS运行过程中的取值为$\{a_i(\mathrm{t}),t=1,2,...,T\}(\underline{x_i}\le a_i(\mathrm{t})\le\overline{x_i})$和$\{a_j(\mathrm{t}),t=1,2,...,T\}(\underline{x_j}\le a_j(\mathrm{t})\le\overline{x_j})$，则它们的可信相关系数r可以用Pearson相关系数定义为：

$$r_{ij}=\frac{\sum_t(a_i(\mathrm{t})-\overline{a_i(\mathrm{t})})(a_j(\mathrm{t})-\overline{a_j(\mathrm{t})})}{\sqrt{\sum_t(a_i(\mathrm{t})-\overline{a_i(\mathrm{t})})^2}\sqrt{\sum_t(a_j(\mathrm{t})-\overline{a_j(\mathrm{t})})^2}}$$

其中，$\overline{a_i(\mathrm{t})}$表示$a_i(\mathrm{t})$的平均值，$\overline{a_i(\mathrm{t})}$表示$a_i(\mathrm{t})$的平均值，r的取值范围在区间[−1，1]内，若可信属性a_i和a_j完全相关，则有r=1或者−1；若可信属性a_i和a_j完全独立，则r=0。

可信属性之间的相关系数是对称的，即对于可信属性a_i和a_j，有$r_{ij}=r_{ji}$。

可信相关系数的绝对值越大，相关性越强——可信相关系数越接近于1或−1，可信属性之间的相关性越强，可信相关系数越接近于0，可信属性之间的相关度越弱。通常情况下可以通过以下的取值范围来判断可信属性自己的相关强度：对于两个可信属性来说，其相关系数的绝对值为0.8—1.0则称两个可信属性之间极强相关，相关系数的绝对值为0.6—0.8称两个可信属性之间强相关，相关系数的绝对值为0.4—0.6称两个可信属性之间中等程度相关，相关系数的绝对值为0.2—0.4称两个可信属性之间弱相关，相关系数的绝对值为0.0—0.2称两个可信属性之间极弱相关或无相关。

可信属性与可信属性之间的独立性和冗余性可以根据可信属性之间的相关性来定义。一般地，如果两个属性是完全

无关的，那么称它们之间是独立的；如果两个属性之间是完全相关的，那么称它们之间是彼此冗余的。

可信属性之间的独立和冗余程度可以定义如下：

如果可信属性a_i，对于任意可信属性$a_j \in A/\{a_i\}$都有$r_{ij}=0$，则称可信属性a_i在可信软件系统TS的可信属性集A中是完全独立的，可信属性a_i也称为可信属性集A中的孤立可信属性。

对于可信属性a_i和a_j有$|r_{ij}|=1$，则称可信属性a_i和a_j完全冗余的，一个属性是完全冗余的应从可信属性集A中删除。对于两个可信属性来说，其相关系数的绝对值大于0.2则称两个可信属性是部分冗余的，部分冗余的可信属性的处理要根据不同可信系统和不同用户的要求来进行，处理时可以根据要求设定一个阈值th，当相关系数的绝对值大于阈值th时看成是冗余的，否则认为是非冗余的。

3.3 可信属性最小子集的求解

在对一个可信软件系统TS进行基于可信属性集的可信性度量时，如何选择合适的可信属性集至关重要。为了能够提高效率和节约资料，可以在全部可信属性集相关性分析的基础上，根据可信系统特征，求出最小可信属性集作为可信度量的依据。

基于可信属性的相关性分析，最小可信属性集的选择如下：

(1)根据可信系统特征选择一个合适的冗余性度量阈值th；

(2)基于冗余性度量阈值th和可信属性之间的相关系数来对全部可信属性集进行分类，把它分为一系列子集，凡是相关系数大于阈值th的可信属性分在一个子集中；

(3)完全独立的可信属性即孤立可信属性应该加入到最小可信属性集中；

(4)如果两个或者多个子集中有一个及以上相同的可信属性，则选择一个可信属性加入到最小可信属性集中；

(5)对于剩余的子集，在每个子集中选择一个可信属性加入到最小可信属性集中。

当然，随着可信软件系统运行环境和用户要求的变更，最小可信属性集也会发生变化，因此需要根据可信属性之间的相关性的变化做出相应的调整和处理，以适应环境和用户的需要。

4 结论(Conclusion)

如果一个软件的行为总是与预期一致，则称该软件可信。软件可信性是软件质量的一种特殊的表现形式，它所关注的是使用层面的综合化的质量属性及其保障形式，涉及多个质量属性的集合以及这些属性的综合与平衡。一般来说，软件的行为及其产生的结果可以通过一组适当的属性来描述，即软件可信性可依据能够反映软件某种可信性的一组属性以及用户在这组属性上的预期来共同刻画，这贯穿于软件的整个生命周期。

软件可信性可由软件可信属性集来进行描述和表示，但是人们根据软件系统的不同特征和用户所关心的不同层面提出了很多描述软件可信的属性集。这些属性集中的属性可能不相同，有些是冗余的，有些甚至是相互冲突的，为此，需要考虑这些可信属性之间的相互关系。本文提出了一种可信属性的表示以及相关性度量与分析的方法，给出了可信属性相关性、独立性和冗余性的概念和度量方法。基于可信属性相关性分析，根据不同的软件类别、应用领域和用户所关注的不同方面和指标，可以从所知的属性集中选择合适的属性来作为对软件可信性进行评估的指标，建立一个适合的可信属性最小子集作为软件可信性度量的体系，并根据环境和用户的需要对可信属性最小子集进行不断地调整和优化。

参考文献(References)

[1] Laprie J C.Dependability:Basic concepts and terminology[M].Vienna:Springe-Verlag,1991.

[2] Algirdas Avizienis,Jean-Claude Laprie,Brian Randell.Fundamental concepts of computer system dependability[J].IARWIEEE·RAS Workshop on Robot Dependability:Technological,Challenge of Dependable Robots in Human Environments.2001,2(15):1-16.

[3] Algirdas A.,et al.Basic concepts and taxonomy of dependable and secure[J].Computing.IEEE Trans.Dependable Secure.2004,1(1):11-33.

[4] NSTC.Research Challenges in High Confidence Systems.In:Proceedings of the Committee on Computing Information and Communications Workshop,1997.

[5] 陈火旺,王戟,董威.高可信软件工程技术[J].电子学报,2003,31(12A):1933-1938.

[6] Steffen Becker,et al.Trustworthy software system:a discussion of basic concepts and terminology[J].ACM SIGSOFT Software Engineering Notes,2006,31(6):1-18.

[7] 刘克,等.“可信软件基础研究”重大研究计划综述[J].中国科学基金,2008(3):145-151.

[8] 郁湧,刘永刚,侯江畔.一种基于构件的可信软件系统框架及其表示[J].软件工程师,2015,18(5):60-62.

[9] Avizienis A,Laprie J C,Randell B.Fundamental concepts of dependability[C].3rd Information Survivability Workshop.

(下转第56页)