

基于软件加权网络的软件结构复杂性度量

田 鹤¹ 赵 海²

(辽宁科技学院工程实践中心 本溪 117004)¹ (东北大学计算机科学与工程学院 沈阳 110004)²

摘 要 软件结构复杂性问题一直备受关注。随着软件规模的不断增大,传统的度量方法已很难适应软件的发展。为了有效地度量软件结构复杂性,以复杂网络理论为基础,分析软件加权网络的强度和紧密度特征参数之间的关系,并统计软件样本集的总体特征量。首先选取特征值较大的软件,研究其对应的软件加权网络中强度和紧密度数值较大的节点。然后对比分析排序差别较大的节点的属性和功能。最后评价两种软件结构复杂性度量方法,得出用强度可以度量软件局部结构的复杂性,紧密度能够度量软件整体结构的复杂性。

关键词 软件度量,加权网络,强度,紧密度,软件结构复杂性

中图法分类号 TP311.5 **文献标识码** A

Metrics for Software Structure Complexity Based on Software Weighted Network

TIAN He¹ ZHAO Hai²

(Engineering Practice Center, Liaoning Institute of Science and Technology, Benxi 117004, China)¹

(School of Computer Science and Engineering, Northeastern University, Shenyang 110004, China)²

Abstract Software structure complexity problems have been closely attention. With the increasing of the software scale, traditional metric methods have been difficult to adapt to the software development. In order to measure the software structure complexity effectively, we analyzed the relationship between the strength and closeness in the software weighted network, and calculated the general characteristics in the software sample collection based on the complex network theory. Firstly, we selected the software with a larger value parameter, and studied the nodes with a larger value in the strength and closeness in its corresponding software weighted network. Secondly, we contrasted and analyzed the attributes and functions of nodes with greater difference in sorting. Finally, we evaluated the two kinds of software structural complexity metric method, and concluded that the strength can measure the local structure complexity of the software, and closeness can measure the whole structure complexity of the software.

Keywords Software metrics, Weighted networks, Strength, Closeness, Software structural complexity

1 引言

在人们的日常生活中,软件的应用领域日益广泛。随着用户需求的变化,软件的规模不断扩大,软件产品更新换代频繁,软件结构越来越复杂^[1],软件的复杂性已经成为软件自身固有的属性。软件开发人员在处理软件复杂性问题方面经常滞后于软件快速发展的步伐,使得问题层出不穷,对软件的维护也造成一定的困难。

互联网的普及推动着软件技术的发展,在软件工程领域中,软件系统复杂性问题的研究得以有效的应用。传统的度量方法只是对软件内的文本、数据及控制信息等软件内部属性进行考量,没有关注软件的宏观结构,这显然已不再适合刻画和度量大规模软件系统结构。

软件网络^[2]的发展已有十余年,在此基础上已累积了大量研究,诸如对软件网络中信息传播的分析^[3]、软件网络中可靠性评估^[4]以及对多粒度软件网络特征的分析^[5]等。随着研究的深入,在软件网络中引入权重能更准确地描述软件结构

和特性^[6,7]。将复杂网络理论引入到软件度量方法中,提出了一个基于软件网络拓扑特征的度量体系,旨在精确地描述软件的结构,同时从宏观角度来度量软件。

本文利用软件加权网络模型中网络拓扑特征值对软件结构进行量化分析,通过对软件样本集的统计特征和单个软件内部结构特征分析,提出用强度和紧密度特征参数来度量软件系统结构的复杂性。评估两种度量方法的可行性在软件工程领域对软件的开发、设计和维护具有指导意义。

2 特征分析

2.1 强度

同复杂网络中的度相似,在加权网络中,用节点强度来表示网络的连通性,节点的强度定义为连接该节点的所有边的权重的和^[8,9]。由于不同边的类型对应不同的权值,节点强度的数值等于所有边权与其边数的乘积,于是强度公式可以表示为:

$$S_i = \sum_{j=1}^N a_{ij} w_{ij} \quad (1)$$

本文受国家自然科学基金项目(60973022)资助。

田 鹤(1985—),女,硕士,主要研究方向为软件网络;赵 海(1959—),男,博士,教授,主要研究方向为复杂网络、软件网络、物联网等。

定义和公式说明了加权网络中强度描述了单个节点的特征,并且相邻节点间的连接与交互反映了环绕某一节点区域的复杂程度。

2.2 紧密度

紧密度是中心化指标之一,用来描述节点在网络中的作用力和影响力,其值定义为该节点到达所有其它节点的距离之和的倒数^[10,11],用公式表示为:

$$C(i) = \left[\sum_{j=1}^N d_{ij} \right]^{-1} \quad (2)$$

其中, d_{ij} 是节点 i 到节点 j 的最短路径长度。

在软件加权网络中,紧密度通常用来衡量节点的中心化程度,一般需要将其归一化^[12],消除网络规模的影响,这样就可以用同一参数在不同的网络规模下进行数值比较。归一化的紧密度公式为^[13]:

$$C_C(i) = (N-1)C(i) \quad (3)$$

与无权网络一样,加权网络中紧密度的数值同样是根据最短路径长度计算出来的,这里最短路径长度的计算过程是带有权值的,所以加权网络的紧密度更能刻画节点的作用方式和波及范围。紧密度参数不仅考虑了节点的复杂度和重用度,还突显了节点在网络中的地位,节点的紧密度越大,说明该节点越重要,对整个网络的影响越大,在网络中处于核心位置^[14]。

2.3 紧密度与强度的关系

强度和紧密度是加权网络中两个重要的拓扑特征参数。本文选取了大量的软件样本,分析了节点强度和紧密度的关系。以 Azureus4.6.02 为例,强度与紧密度的关系如图 1 所示。

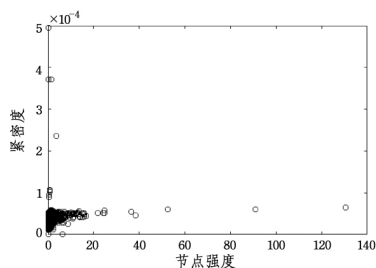


图1 Azureus4.6.02 软件加权网络中强度与紧密度的分布关系

图 1 中横坐标是软件加权网络中节点强度的数值,纵坐标是对应节点紧密度的数值。对横纵坐标取对数可以看出,随着节点强度的增大,大多数节点的紧密度都趋于一个稳定的值,特别地,有一些强度很小的节点的紧密度反而很大,这种现象说明了在软件加权网络中存在极少数重要节点,它们处于核心位置,其它大多数节点的中心化程度大致相同。

3 实证分析

本文将软件抽象成加权网络模型,用强度和紧密度特征参数量度软件结构的复杂性。具体按照如下步骤进行实验。

步骤 1 利用 SPSS 软件对已抽取的软件样本集的各软件加权网络的两个特征参数值进行统计和 K-S 分布检验,了解它们在加权网络中的分布情况。在此过程中,将所选取的软件样本集的平均强度和平均紧密度分别与其节点数作回归分析,得出平均强度与节点数是不相关的,而平均紧密度与节点数是相关的,即平均强度与软件规模无关,平均紧密度与软

件规模有关,这表明了紧密度指标是中心化的指标。将紧密度中心化后即为中心化紧密度,它与软件规模是不相关的,这样才能有效地度量软件。

步骤 2 对每个样本的软件加权网络的平均强度和平均中心化紧密度的数值作统计,由统计数据确定它们大致分布范围。经统计,平均强度的数值主要分布在 $[0.51896, 1.93522]$ 区间内,平均中心化紧密度的数值主要集中在 $[0.000519, 0.151348]$ 范围内。

步骤 3 对平均强度和平均中心化紧密度进行 K-S 正态分布检验,由检验数据知它们都服从正态分布。分别取置信度为 95% 的置信区间,对超出取值范围很多且数值很大的平均特征参数来说,正是因为软件加权网络中大部分节点的特征参数值较大才使得网络整体平均参数值很大,同时有理由认为这些软件结构是比较复杂的。于是选出这样的软件来作进一步分析。

平均强度最大的软件是 Abiword2.4.6。软件加权网络中节点强度的分布如图 2 所示。

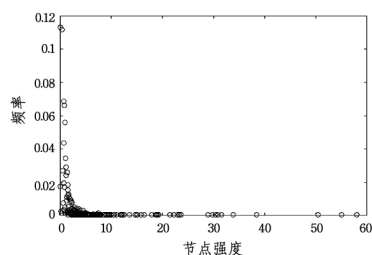


图2 Abiword2.4.6 软件加权网络中节点强度的频率分布

从图中可以明显看到有 3 个节点强度值非常大,查找节点相关信息,选择最右边的两个节点进行分析。强度最大的是第 1621 号节点,名称是 `UT_GenericVector`,它是一个通用向量的类,可以存放动态数组,具有多种数据结构,适用于多种算法。`UT_String` 节点强度的数值也很大,它是 1662 号节点,是一个字符串类,程序中对字符串的定义都要用这个类,对字符串的基本操作也都是针对 `UT_String` 类型进行操作。从这些节点的性质可以看出,它们都是一些基础类,实现一些简单的基本功能,在软件结构中通常位于底层,其它的类模块直接调用它们,因此这些类与其它类模块直接作用关系比较密切,所以强度值很大。在软件开发过程中,调整软件内部结构的布局,可以降低这些类模块间的复杂性,使整个软件结构不会过于复杂。

平均中心化紧密度最大的软件为 openssl-0.9.7l,这是一个强大的安全套接字层密码库。分析其软件加权网络内的节点中心化紧密度,频率分布如图 3 所示。

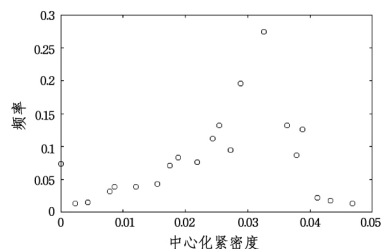


图3 openssl-0.9.7l 软件加权网络中节点中心化紧密度的频率分布

图中的分布呈陡峭的单峰形,数值最大的是第 37 号节点,名称是 `asn1_string_st`,其功能是定义字符数据,其属性包

括长度、标记等。第 81 号名为 `crypto_ex_data_st` 的节点的中心化紧密度也很大,它是一个空栈的类,里面只有一个整型数据的实体,类结构简单。堆栈时通常要先把指针指向下一个空位置,用来存放数据,这时就要调用这个类。这些中心化紧密度的数值非常大的节点都是软件中比较重要的模块,其它类模块都必须调用这些模块才能完成对数据结构的基本操作以及源程序的执行。同时,对整个软件来说,实现软件的主要功能都需要把这些类模块调用起来,因此它们对于软件的整体结构来说是非常重要的,是核心模块。

4 两种度量方法的评价

通过对节点强度和紧密度的属性及特征关系研究得出,在软件加权网络中,大部分节点紧密度的数值都非常接近,这样的稳态并没有因为节点强度的增大而改变。只有几个节点很特殊,它们的紧密度非常大,相应的强度却比较小。依据这样的网络宏观拓扑特性来分析软件内部结构的复杂性。以 `Mysql6.0.6` 为例,将这个软件加权网络的节点强度和紧密度分别排序,选出几个特例,将排序结果相差很多的节点列于表中,具体如表 1 所列。

表 1 `Mysql6.0.6` 软件加权网络中节点强度与紧密度的排序比较

紧密度排序	强度排序	节点序号	节点名称
1890	6	432	<code>Create_func_arg1</code>
584	10	1790	<code>Item_int_func</code>
1529	15	2746	<code>SerialLogControl</code>
2545	18	853	<code>debug_type</code>
3	180	423	<code>Create_field</code>
8	242	2146	<code>Ndb_cond_traverse_context</code>
14	349	1816	<code>Item_param</code>
25	737	3023	<code>st_ha_create_information</code>

选取两个节点进行研究,如 `Create_func_arg1`,这个类能够动态地创建函数的参数,如果创建子类,可以分别设置函数参数的个数。对于软件中单个模块的开发来说,函数是程序的基本组成部分,把函数的创建及参数的定义放在一个类模块中,便于其它模块随时能直接调用,而这样的交互是随时发生的,会使软件内部局部构造复杂,这就是节点强度大的原因。而从另一个角度来说,创建函数也可以在类模块内部完成,并不一定要调用一个基类,对于整个软件加权网络拓扑结构来说,它的影响力很小,所以节点紧密度的数值很小。再如第 2146 号节点,`Ndb_cond_traverse_context` 是一种存储引擎类,根据条件数来遍历上下文。在对数据库操作时经常在后台运行,承担着对软件操作过程中核心功能的实现,相应的在软件加权网络中处于重要位置,并且是复杂节点,如果对此节点修改将会波及整个网络,说明它的紧密度很大。但是与它直接相连的节点并不多,出于对模块功能的考量很少有类模块直接调用,所以它的局部结构比较简单,节点强度并不大。

通过实例分析,软件加权网络中的强度和紧密度特征参数都能描述软件结构的特征,度量软件结构的复杂性。在局部范围内,强度表达了软件内模块间的连接和相互作用,说明各类模块的复杂性。而紧密度指标体现了节点在整个网络的影响力,相应地说明了软件内各模块在整个软件中的重要程

度,从软件的功能和应用方面来分析模块的复杂性,进而度量整个软件结构的复杂性。这样从局部和整体都能刻画软件结构的复杂性,在实际度量软件时可以将两种方法相结合,或者根据现实需要选择其中一种度量方法。

结束语 软件的度量贯穿于软件生命周期的各个阶段,目的是提高软件质量,使软件能够适应计算机技术的发展。在此过程中软件会出现各种漏洞和缺陷,其中软件结构复杂性问题是重要方面。如何能够高效而准确地度量软件结构复杂性是软件维护中亟待解决的问题。

本文通过软件加权网络拓扑特征结构,首先分析了软件加权网络中强度和紧密度特征参数的属性及相互关系,为度量软件结构复杂性提供了理论依据。接着对软件样本集中的软件总体特征作统计分析,结合单样本软件特征参数的分布规律,得到强度,其体现了以某一节点为中心的局部复杂程度,可以从局部来度量软件结构的复杂性,紧密度指标体现了节点在整个网络中的影响力,可以从全局的角度来度量软件结构复杂性。在软件开发、设计和维护过程中,根据实际情况和现实需求可以选择其中一种度量方法,也可以两种方法并用来全面度量软件结构的复杂性。

参 考 文 献

- [1] Cai Kai-yuan, Yin Bei-bei. Software Execution Processes as an Evolving Complex Network[J]. Information Sciences, 2009, 179 (12): 1903-1928
- [2] 何克清, 马于涛, 刘婧, 等. 软件网络[M]. 北京: 科学出版社, 2008: 1-117
- [3] 吴逸伦, 张博锋, 赖志权, 等. 基于消息语义解析的软件网络行为分析[J]. 计算机应用, 2012, 32(1): 25-29
- [4] 刘铮. 大规模软件可信性度量分析原理[D]. 沈阳: 东北大学, 2011
- [5] 韩言妮, 李德毅, 陈桂生. 软件网络的多粒度拓扑特性分析及其应用[J]. 计算机学报, 2009, 32(9): 1712-1721
- [6] 张志华, 封筠, 赵正旭. 工程软件格式加权网络的复杂特征分析[J]. 计算机工程与应用, 2011, 47(31): 82-85
- [7] Myers C R. Software Systems as Complex Networks; Structure, Function and Evolvability of Software Collaboration Graphs[J]. Physical Review E 68. 2003, 68(4)
- [8] 汪北阳, 吕金虎. 复杂软件系统的软件网络结点影响分析[J]. 软件学报, 2013, 24(12): 2814-2829
- [9] 孙兵. 加权网络演化模型的研究[D]. 大连: 大连海事大学, 2009
- [10] 王林, 张婧婧. 复杂网络的中心化[J]. 复杂系统与复杂性科学, 2006, 3(1): 13-20
- [11] 李茂林, 龙建国, 张德群. 基于复杂网络理论的作战体系节点重要性分析[J]. 指挥控制与仿真, 2010, 32(3): 15-17
- [12] 张慧. 多粒度软件网络的特性研究与演化分析[D]. 沈阳: 沈阳工业大学, 2013
- [13] 司晓静. 复杂网络中节点重要性排序的研究[D]. 西安: 西安电子科技大学, 2012
- [14] 王文钊, 汪斌强, 王志明, 等. 基于网络中心性分析的虚拟网络映射算法[J]. 计算机应用研究, 2015, 32(2): 565-568