

基于复杂网络的软件结构度量方法综述

孙世温^{1,2}, 夏承遗^{1,2}, 王莉^{1,2}

(1. 天津理工大学 计算机与通信工程学院, 天津 300191; 2. 天津理工大学 天津市智能计算及软件新技术重点实验室, 天津 300191)

摘要: 计算机软件复杂性与软件质量、开发成本和生产效率等密切相关, 软件复杂性的度量和控制是计算机科学的挑战性问题之一. 近年来复杂网络研究的兴起为研究软件系统结构复杂性提供了新的理论、方法和工具, 该方法克服了传统软件结构度量方法侧重微观统计、缺乏全局性和整体性等缺点, 构成了复杂网络与传统软件工程的交叉研究领域. 对该领域的研究工作进展进行了介绍, 从软件网络结构特征分析、建模以及研究成果的初步应用 3 个方面总结已有工作, 并对今后有意义的研究方向进行了展望, 如基于加权模型的软件网络建模、软件网络动态演化机制等.

关键词: 复杂网络; 软件结构复杂性; 软件度量; 建模

中图分类号: TP311.5 **文献标识码:** A **文章编号:** 1673-4785(2011)03-0208-05

Survey of the measurement of software structures based on complex networks

SUN Shiwen^{1,2}, XIA Chengyi^{1,2}, WANG Li^{1,2}

(1. School of Computer and Communication Engineering, Tianjin University of Technology, Tianjin 300191, China; 2. Tianjin Key Laboratory of Intelligence Computing and Novel Software Technology, Tianjin University of Technology, Tianjin 300191, China)

Abstract: Software complexity is closely related to software quality, development cost, and development efficiency. Thus the measurement and control of software structural complexity is one of the most challenging problems in computer science. Recently, the research of complex networks has been providing new theories, methods, and tools for the study of software structural complexity, and overcoming some shortcomings of traditional measurement methods, such as only focusing on the microstructures and a lack of structural integrality, which results in the interdisciplinary field of complex network and traditional software engineering. In this paper, the following aspects of the research work in this new area were summarized and reviewed: topological structure analysis, network modeling, and the application of complexity controls and structural optimization. A perspective for meaningful future research emphasis was given including aspects such as network modeling based on weighted models and dynamical evolving mechanism of software systems.

Keywords: complex network; software structural complexity; software metrics; modeling

现代计算机软件系统的复杂程度越来越高, 越复杂的软件就越难以保证其产品质量、开发成本和生产效率, 软件开发极易处于失控状态, 甚至形成“雪崩

效应”. 大型软件系统由无数细节构成, 一个极小的错误就可能引发灾难性后果. 因此, 如何认识、度量、管理和控制软件复杂性已成为软件工程领域中极其重要且极具挑战性的问题. 对软件结构复杂性度量和评估的前提是需要对软件结构(信息)进行合理描述和有效量化^[1-2]. 传统软件度量方法^[3-5]注重从单一模块出发, 忽略了软件系统结构的整体性, 侧重微观上的统计, 缺乏全局和整体上对软件结构的度量; 而计算机软件作为一类复杂系统, 系统(拓扑)结构必然会影响其功能、性能和质量等. 近年来, 复杂网络研究取

收稿日期: 2010-03-30.

基金项目: 国家自然科学基金资助项目(60904063); 天津市应用基础及前沿技术研究计划资助项目(11JCYBJC06600); 中国博士后研究基金资助项目(20090460694); 天津市高等学校科技发展基金资助项目(20090717, 20090811, 20090813); 国家大学生创新性实验计划资助项目(091006007, 101006019).

通信作者: 孙世温. E-mail: sunsw80@gmail.com.

得了相当大的进展^[6-10],其理论、技术、方法和成果为从整体和全局角度研究软件系统结构和行为复杂性提供了新的工具。

按照软件工程的观点,在不同粒度上,计算机软件是由多个功能实体(类、对象、方法、函数、子程序等)相互关联、共同作用构成的一类复杂系统;从网络化的角度,实体映射为节点,实体间存在某种联系映射为相应节点的连接边,软件系统的结构可以通过网络模型进行描述和刻画。应用复杂网络研究领域的理论、技术和方法,从整体和全局的角度来探索软件系统的结构和行为特性,弥补了传统软件度量技术的欠缺,构成了复杂网络和软件工程的交叉研究领域(以下简称为“复杂软件网络”)。

本文着眼于复杂网络方法在软件结构复杂性度量方面的应用,在概述和分析几种传统度量方法的基础上,从软件系统网络化特征分析、软件网络建模、研究成果应用3个方面简要阐述了国内外在复杂软件网络领域的一些主要研究工作,并尝试对该领域的研究热点和难点进行了归纳和总结。

1 传统的软件结构度量方法

传统软件结构度量方法包括基于控制流的 McCabe 环复杂度度量^[3]、基于程序体积的 Halstead 方法、面向对象软件的 C&K 度量^[4]、MOOD 方法^[5]等。

1.1 McCabe 环复杂度

McCabe 度量方法^[3]根据图论和程序结构控制理论,把程序结构的控制流程图转化为有向图(即程序图),然后将强连通有向图的环数作为程序结构的环形复杂度 $V(G) = m - n + 1$,其中 m 和 n 分别为有向图 G 中的弧数和节点数, $V(G)$ 还等于程序图中判定结点的数目加 1。

McCabe 度量方法的计算过程表明程序复杂性很大程度上取决于程序结构控制流的复杂性。环形复杂度高的程序往往是最容易出现错误的程序,实践表明程序规模以 $V(G) \leq 10$ 为宜。

1.2 Halstead 方法

Halstead 方法通过计算程序中运算符和操作数的数目对程序复杂性加以度量。设 n_1 为程序中不同运算符的数目, n_2 为程序中不同操作数数目, N_1 为程序中运算符总数, N_2 为程序中操作符总数,令 H 和 N 分别表示程序的预测长度和实际长度, Halstead 给出计算公式:

$$H = n_1 \times \lg n_1 + n_2 \times \lg n_2,$$

$$N = N_1 + N_2.$$

Halstead 的重要结论之一是程序的实际长度与预测长度非常接近,即使程序还未编写完成也能估算出程序的实际长度。Halstead 方法还给出了另外

一些计算公式,包括:程序容量 $V = N \times \lg(n_1 + n_2)$ 、程序级别 $L = (2/n_1)(n_2/N_2)$ 、编制程序所用工作量 $E = V/L$ 、程序中错误数预测值 $B = N \times \lg(n_1 + n_2)/3\ 000$ 等。

Halstead 度量实际上只考虑了程序的数据流而没有考虑程序的控制流,因而不能从根本上反映出程序的复杂性。

1.3 C&K 度量

1994 年 Chidanber 和 Kemerer 阐述了面向对象软件度量学的理论基础,提出了基于继承树的面向对象度量方法(简称为 C&K 度量方法)^[4],包括 6 个度量指标:1)类的加权方法数;2)继承树的深度,若为多重继承则计算从节点到树根的最大深度;3)子类数目,即一个类的直接子类数目;4)对象间耦合程度,当一个类使用其他类的成员变量或方法时,则称这 2 个类是耦合的;5)类的响应个数,即为类中所有方法所调用的类外方法的总数目;6)类方法中内聚缺乏度。

1.4 MOOD 方法

1995 年 Brito 提出的 MOOD 方法间接度量了面向对象软件的封装性、继承性、耦合性和多态性^[5],包括:1)封装性:方法隐藏因子、属性隐藏因子分别度量隐含方法数、属性数相对于方法总数、属性总数的比例;2)继承性:方法继承因子和属性继承因子分别度量直接继承的方法数、属性数相对于方法总数、属性总数的比例;3)耦合性:耦合因子是对系统中所有成类对间的关联程度大小的度量;4)多态性:多态性因子表示重载方法数相对于新方法总数的比例。

上述几种传统度量方法从不同侧面描述了软件结构的复杂性,但偏重于分析软件系统中功能个体(类、过程、方法等)的性质和局部结构,因而缺乏全局和整体上对软件结构的度量。

2 复杂网络概述

20 世纪 90 年代以来,借助于计算机运算能力的增强以及数据存储、处理技术的快速发展,科学家将现实世界中复杂系统的结构进行网络化处理,将系统中的个体抽象为网络的节点,个体间存在的相互联系和作用视为节点间存在连边。研究发现许多真实存在的复杂系统都具有,以小的节点间距和高的节点聚集程度为特征的小世界效应(small-world)^[6]和以节点度分布服从幂律为特征的无标度特性(scale-free)^[7]。小世界和无标度的发现促成了复杂网络这一新的研究领域的产生,此后物理、生物、数学、计算机、经济、管理、金融学等不同学科领域的研究者都着重关注复杂系统的拓扑连接结构与

系统功能及动力学性质之间的相互作用和影响^[8-10]。

3 基于复杂网络的软件结构

基于复杂网络的软件结构模型用形式化的方法来描述软件系统中实体之间的关系,具有良好的数学基础,为从系统整体上分析软件结构的复杂性提供了新的思路,能够得到由软件系统中单个个体的性质累加所无法得到的整体性质。国内外研究者在复杂软件网络领域的研究工作大致可以分为3类,分别是软件系统网络化结构特征分析、基于建模的网络特征形成机理和演化规律研究,以及研究成果在软件复杂性控制、结构优化等方面的应用。

3.1 软件系统网络化结构特征分析

Valverde 等人^[11]将面向对象软件中类及类间的继承、关联等关系映射为无向网络,发现类网络表现出无标度和小世界特征,无标度特性表明软件系统中各个模块间不是毫无规律地随机连接,系统中存在少量连接度极高的类,因而这些类耦合程度较高,小世界特征表明节点平均间距较小且节点聚集程度较高。通过赋予边方向特性,Myers^[12]基于有向网络模型对类网络进行连通特性、出度入度匹配关系、层次结构等方面的深入研究,发现节点出度和入度分布都服从幂律,且两者具有负相关性,即出度小的节点通常入度较大,文献从“模块复用”角度给出了解释:出度小的类很少依赖其他类,较为简单且被重用的次数多的类其入度就可能较大。

Shi 等人^[13]首次分析了类网络的演化性,即随着开发时间或版本变更软件系统中网络结构特征的变化规律。韩明畅等人^[14]、陈焘等人^[15]对 Java 类库中类以及类关系的研究,进一步验证了类网络中小世界和无标度规律的普适性。Ma 等人^[16]提出了一个复合度量模型,从代码级、类级和系统级3个层面给出面向对象软件系统的度量。Concas 等人^[17]发现面向对象属性分布符合幂律,且网络结构存在自相似、分形等特点^[18]。

此外,Challet 等人^[19]利用有向网络,描述了 Redhat 8.0 源代码中包的关联关系和函数间调用关系,统计了网络入度和出度分布、出度入度关联等特征,并最早将复杂网络传播动力学的思想引入到软件缺陷传播研究中。Lai 等人^[20]采用无权网络刻画了开源软件中源代码文件间的引用关系,发现了软件网络中的小世界和无标度特征。笔者^[21-22]对 Linux kernel 源代码进行解析,提出一种基于依赖关系密切程度的权值表征方法,构建了2类加权网络模型(HFCN-I 和 HFCN-II)来刻画功能模块(头文件)的依赖关系,发现网络具有小世界、连接极不均

匀、度分布“胖尾”、权分布服从幂律、“强强联合”效应等特征。

3.2 软件网络建模研究

软件网络建模是在分析真实网络特征和捕捉网络特征形成机制基础上,建立数学模型和算法,揭示软件系统中网络特征的形成机理和演化规律。

复杂网络领域中建模研究成果对构建软件网络演化模型具有借鉴意义。Watts 和 Strogatz 提出了著名的小世界模型^[6],通过在规则网络中按照一定概率添加长程边解释了“小世界”的产生原因;Barabási 和 Albert 在揭示 Internet 和万维网中度分布服从幂律这一重大发现后,提出了著名的无标度模型^[7],揭示无标度特性的出现取决于2点——网络的动态增长(新节点加入)和边择优连接,即新加入的节点依概率倾向于与连接边多的节点相连。在 BA 模型的基础上研究人员对无标度的产生机制进了一系列深入探索,如考虑初始吸引度、非线性择优连接、择优连接更迭、增长制约条件、增长方式及局部作用等^[8-10]。

考虑软件网络的演化特性,文献[20, 23]认为软件网络中无标度特性的出现也归因于软件本身的进化及进化过程中存在的“择优连接”机制,即软件网络中存在着关键个体(模块、类、对象、函数等),新个体添加到系统时往往倾向于和这些关键个体连接。

Valverde 等人^[24]认为全局性约束和设计优化,即以较小代价(较短的平均最短路径)获得可靠的节点间通信是软件系统展现出“小世界”特性的主要因素,在此基础上,Valverde 等人还提出了一个基于复制和分岔规则的软件演化模型^[25],发现软件网络的生长遵循某种对数规律(如系统的总边数、平均度随时间变化呈对数增长)。

Myers 等人^[12]认为软件系统中的优先连接实质上是一种软件重用机制,节点的入度越大表明其重用率越高,通过有效的重用能形成度很大的中枢节点,从而有利于节点间的通信。He 等人^[26]提出了基于模式冷冻的生长算法,为研究软件系统的演化规律提供了新的思路和方法。李兵等人^[27]则提出了一种生长的网络演化模型 CN-EM,借鉴了遗传算法中的交叉、变异等演化算法,通过对比性分析显示模型网络等能较好地刻画实际软件系统中复杂网络特性出现的演化过程。

3.3 研究成果应用

将软件网络特征分析和度量结果付诸软件工程实践,探讨在软件复杂性控制、质量预测、结构设计优化等方面的应用研究也得到不少学者的关注。Vassa 等人^[28-29]根据网络边数和节点数之间的关系研究结构稳定性变化,从而预测软件规模和开发成本;

Girolamo 等人^[30]根据介数等指标在类层、网络层和设计层中识别和检测软件结构缺陷以及隐患类,进行软件质量预测;Melton 等人^[31]发现类间形成依赖环会导致系统复杂度增加、稳定性降低;Ma 等人^[32]发现网络中重要子图结构都较稳定且没有环出现,从2个不同的角度验证了“避免环产生”这一结构设计准则的合理性。

针对软件复杂性问题,基于复杂网络的研究方法及目前取得的研究成果不仅丰富了对软件系统结构复杂性特征的刻画,更为控制软件系统复杂性,指导软件结构设计等提供了一个新的研究途径,对于复杂软件体系结构优化、软件质量提高等方面都具有重要的现实意义。

4 结 论

目前,复杂软件网络领域的研究还处于初期探索阶段,要最终形成成熟的应用技术还需要大量更具创新性的细致和完善的工作。展望未来,复杂网络理论和方法在计算机软件系统中的研究和应用需要更好地结合软件系统本身的特点,建立更适合软件系统的网络模型。

1) 基于网络模型的软件系统拓扑结构分析发现了传统软件度量方法所忽略的结构特征,但由于上述研究均采用无权网络,不能体现软件系统中个体间交互作用的强度与差异对系统的影响;因此将度量要素以权重的形式赋予节点和边,建立加权网络模型,将会更加全面细致地刻画软件系统的结构和行为。另外,如何将传统的软件度量方法和基于网络模型的软件结构分析相结合,并分析不同度量指标之间的内在联系等问题也有待深入研究。

2) 已有软件网络建模研究借鉴了一般复杂网络的演化机制(如“节点新增”、“边择优连接”等),探求了软件网络中静态拓扑结构特征(小世界、无标度等)的形成机理;但软件作为一种人工设计和开发的智能化产物,软件设计开发技术和方法(如面向对象原则、代码重构技术等)的应用是软件系统结构自身不断演化的重要原因之一,而现有的演化模型在这方面没有深入考虑。由此在进一步的研究中要着眼于分析软件设计开发技术和方法对软件系统局部结构的作用,探索局部规则作用下如何产生软件网络整体特征的涌现。

在“软件的网络观”得到计算机科学领域研究者认可和日渐重视的背景下,笔者有理由坚信复杂网络与计算机软件工程的交叉研究将有助于更加客观全面地定量评估软件结构,在软件产品的设计、开发、测试和维护等阶段,能够为实施复杂性控制、降

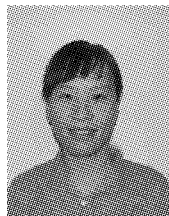
低开发风险和成本、进行早期软件质量预测、制定正确测试策略等提供科学的理论依据。总之,基于复杂网络的软件结构复杂性研究是一个非常具有实用价值的研究方向,并且有待进一步深入。

参考文献:

- [1] 杨美清. 软件工程技术发展思索[J]. 软件学报, 2005, 16(1): 1-7.
YANG Fuqing. Thinking on the development of software engineering technology[J]. Journal of Software, 2005, 16(1): 1-7.
- [2] NORMAN E F, SHARI L P. Software metrics[M]. 2nd ed. Beijing: China Machine Press, 2003: 5-10.
- [3] MCCABE T J. A complexity measurement[J]. IEEE Transaction on Software Engineering, 1976, 2(4): 302-308.
- [4] CHIDAMER S R, KEMERER C F. A metrics suite for object-oriented design[J]. IEEE Transactions on Software Engineering, 1994, 20(6): 476-493.
- [5] ABREU F B E. The MOOD metrics set[C]//Proceedings of ECOOP'95 Workshop on Metrics. Aarhus, Denmark, 1995: 150-152.
- [6] WATTS D J, STROGATZ S H. Collective dynamics of small world networks[J]. Nature, 1998, 393(6684): 440-442.
- [7] BARABASI A L, ALBERT R. Emergence of scaling in random networks[J]. Science, 1999, 286(5439): 509-512.
- [8] BOCCALETIA S, LATORAB V, MORENOD Y, et al. Complex networks: structure and dynamics[J]. Physics Reports, 2006, 424(4/5): 175-308.
- [9] WANG Xiaofan, CHEN Guanrong. Complex networks: small-world scale-free and beyond[J]. IEEE Circuit System Magazine, 2003, 3(1): 6-20.
- [10] 陈关荣. 复杂网络及其新近研究进展简介[J]. 力学进展, 2008, 38(6): 653-662.
CHEN Guanrong. Introduction to complex networks and their recent advances[J]. Advances in Mechanics, 2008, 38(6): 653-662.
- [11] VALERDE S, CANCHO R F, SOL R V. Scale-free networks from optimal design[J]. Europhysics Letters, 2002, 60(4): 512-517.
- [12] MYERS C. Software systems as complex networks: structure, function, and evolvability of software collaboration graphs[J]. Physical Review E, 2003, 68(4): 046116.
- [13] SHI Mingjiang, LI Xiang, WANG Xiaofan. Evolving topology of Java networks[C]//Proceedings of 6th World Congress Control on and Automation. Dalian, China, 2006, 1: 21-23.
- [14] 韩明畅, 李德毅, 刘常昱, 等. 软件中的网络化特征及其对软件质量的贡献[J]. 计算机工程与应用, 2006(20): 29-31.
HAN Mingchang, LI Deyi, LIU Changyu, et al. Networked characteristic in software and its contribution to

- software quality[J]. Computer Engineering and Application, 2006(20): 29-31.
- [15] 陈焘, 李孔文, 王树森, 等. 基于复杂网络的 Java 程序分析工具设计与实现[J]. 计算机科学, 2009, 36(4): 145-150.
- CHEN Tao, LI Kongwen, WANG Shusen, et al. Design and implementation of a tool of Java program analysis based on complex networks[J]. Computer Science, 2009, 36(4): 145-150.
- [16] MA Yutao, HE Keqing, DU Dehui, et al. A complexity metrics set for large-scale object-oriented software systems [C]//Proceedings of 6th IEEE International Conference on Computer and Information Technology. Seoul, Korea, 2006: 189-194.
- [17] CONCAS G, MARCHESI M, PINNA S, et al. On the suitability of Yule process to stochastically model some properties of object-oriented systems [J]. Physica A, 2006, 370(2): 817-831.
- [18] CONCAS G, LOCCI M F, MARCHESI M, et al. Fractal dimension in software networks[J]. Europhysics Letters, 2006, 76(6): 1221-1227.
- [19] CHALLET D, LOMBARDONI A. Bug propagation and debugging in asymmetric software structures [J]. Physical Review E, 2004, 70(4): 046109.
- [20] MOURA A, LAI Yingcheng, MOTTER A E. Signatures of small world and scale-free properties in large computer programs[J]. Physical Review E, 2003, 68(1): 017102.
- [21] SUN Shiwen, LIU Zhongxin, CHEN Zengqiang, et al. Header file collaboration networks: weight, topology and statistical properties [J]. Dynamics of Continuous, Discrete and Impulsive Systems Series B: Applications & Algorithms, 2007, 14(S6): 142-147.
- [22] SUN Shiwen, XIA Chengyi, CHEN Zhenhai, et al. Generalized collaboration networks in software systems: a case study of Linux kernels[J]. Frontiers of Computer Science in China, 2009, 3(3): 421-426.
- [23] 闫栋, 祁国宁. 大规模软件系统的无标度特性与演化模型[J]. 物理学报, 2006, 55(8): 3799-3804.
- YAN Dong, QI Guoning. The scale-free feature and evolving model of large-scale software systems[J]. Acta Physica Sinica, 2006, 55(8): 3799-3804.
- [24] VALVERDE S, SOLE R V. Hierarchical small worlds in software architecture [EB/OL]. [2010-03-15]. <http://arxiv.org/abs/cond-mat/0307278>.
- [25] VALVERDE S, SOLE R V. Logarithmic growth dynamics in software networks[J]. Europhysics Letters, 2005, 72(5): 858-864.
- [26] HE Keqing, PENG Rong, LIU Jing, et al. Design methodology of networked software evolution growth based on software patterns[J]. Journal of Systems Science and Complexity, 2006, 19(2): 157-181.
- [27] 李兵, 王浩, 李增扬, 等. 基于复杂网络的软件复杂性度量研究[J]. 电子学报, 2006, 34(12A): 2371-2375.
- LI Bing, WANG Hao, LI Zengyang, et al. Software complexity metrics based on complex networks[J]. Acta Electronica Sinica, 2006, 34(12A): 2371-2375.
- [28] VASA R, SCHNEIDER J G, WOODWARD C, et al. Detecting structural changes in object oriented software systems [C]//Proceedings of International Symposium on Empirical Software Engineering. Noosa Heads, Australia, 2005: 479-486.
- [29] VASA R, SCHNEIDER J G, NIERSTRASZ O. The inevitable stability of software change [C]//Proceedings of 23rd IEEE International Conference on Software Maintenance. Paris, France: IEEE Press, 2007: 4-13.
- [30] GIROLAMO A, NEWMAN L I, RAO R. The structure and behavior of class networks in object-oriented software design [EB/OL]. [2010-03-15]. www.eecs.umich.edu/eenewm/documents/classnetworks.
- [31] MELTON H, TEMPERO E. Static members and cycles in Java software [C]//Proceedings of 1st International Symposium on Empirical Software Engineering and Measurement. Madrid, Spain, 2007: 136-145.
- [32] MA Yutao, HE Keqing, LIU Jing. Network motifs in object-oriented software systems [J]. Dynamics of Continuous, Discrete and Impulsive Systems Series B: Applications and Algorithms, 2007, 14(S6): 166-172.

作者简介:



孙世温, 女, 1980 年生, 讲师, 博士, 主要研究方向为复杂动态网络、网络鲁棒性、软件工程。目前主持天津市高等学校科技发展基金项目 1 项。发表学术论文近 10 篇, 其中被 SCI 检索 3 篇、EI 检索 4 篇。



夏承遗, 男, 1976 年生, 副教授, 博士, 主要研究方向为复杂系统与复杂网络建模与分析、传播动力学。目前主持国家自然科学基金项目 1 项、天津市应用基础及前沿技术研究计划项目 1 项、中国博士后研究基金项目 1 项、天津市高等学校科技发展基金项目 1 项。发表学术论文 10 余篇, 其中被 SCI 检索 3 篇、EI 检索 6 篇。



王莉, 女, 1979 年生, 讲师, 博士, 主要研究方向为复杂网络、多智能体系统的协调与控制。目前主持天津市高等学校科技发展基金项目 1 项。发表学术论文近 10 篇, 其中被 SCI 检索 1 篇、EI 检索 3 篇。