

基于用例图的软件系统复杂网络特性度量

秦怀斌^{1,2}, 郭 理²

(1 中国农业大学 信息与电气工程学院, 北京 100193;

2 石河子大学 信息科学与技术学院, 新疆 石河子 832003)

摘 要: 近年来迅速发展的复杂网络理论为解决软件系统复杂性度量提供了一个新的视角. 同时, 鉴于用例图即是从系统整体角度及整体运行过程出发描述执行者与系统提供的用例之间的连接执行关系, 提出一种将用例图和复杂网络理论相结合进行软件系统复杂网络特性度量的方法. 给出了具体度量过程和算法, 并对提出的度量方法通过实例进行验证.

关键词: 用例图; 复杂网络; 有向加权网络; 度量

中图分类号: TP311

文献标识码: A

文章编号: 1000-7180(2012)07-0072-04

Metrics for Software Systems Complex Networks Characteristics Based on Use Case Diagram

QIN Huai-bin^{1,2}, GUO Li²

(1 College of Information and Electrical Engineering, China Agricultural University, Beijing 100193, China

2 College of Information Science and Technology, Shihezi University, Shihezi 832003, China)

Abstract: The theory of complex networks recently studied has offered a new perspective on solving the problem. Meanwhile, as the use case diagram is start from the system overall perspective and overall operating process that describe the executive and the system to provide the connection between the cases to execute relations, puts forward the complex network characteristics metrics method of software system based on use case diagram and complex network theory, and gives a specific metrics process and algorithm. Finally, the algorithm is verified by example.

Key words: use case diagram; complex networks; directed weighted networks; metrics

1 引言

随着计算机软件系统规模逐渐增大, 其复杂性明显增加. 对软件系统整体结构和整体运行过程的把握及了解程度, 直接影响软件系统复杂性的准确度量, 进而对整个软件系统的设计、实现、集成、测试、维护等阶段产生重要影响.

统一建模语言 UML (unified modeling language) 的用例图即是从系统整体角度及整体运行过程出发, 用于表达若干执行者 (actor) 与用例之间的连接执行关系. 从外部用户视角出发, 一个用例是执

行者与目标软件系统之间一次典型的交互; 从软件系统内部视角出发, 一个用例代表系统执行的一系列动作, 动作执行的结果能被外部执行者所察觉.

复杂网络作为复杂性科学研究的一部分, 致力于复杂系统非线性和非均衡方面的研究. 当一个系统由相互作用的大量构件构成, 其整体的复杂性超过了对构成构件行为的简单叠加, 这个系统就是复杂系统. 迄今为止, 复杂网络特性的研究主要集中在无向无权网络上, 即网络节点的关系是布尔关系, 且为双向关系, 如果节点之间有关联则连接强度值为 1, 节点之间无关联则连接强度值为 0.

收稿日期: 2011-12-15; **修回日期:** 2012-01-09

基金项目: 国家自然科学基金 (60963006); 教育部人文社科项目 (11XJJAZH001); 石河大学科学技术研究发展计划 (ZRKX2010YB18)

通过理论分析及大量实践表明,真实软件系统构件之间的调用及依赖关系有明显的方向性,且系统构件之间的关系有强弱之分,采用单纯的无向无权网络(布尔网络)很难精确地描述真实系统的构成构件、系统运行过程及系统整体结构的复杂特性。鉴于此,论文拟基于用例图对软件系统复杂网络特性进行度量,首先通过用例图对系统整体运行过程进行分析描述,然后用有向加权复杂网络对软件系统复杂网络特性进行描述及度量,以此从较高层次精确度量软件系统的复杂性,为系统后续开发及维护提供指导。

2 复杂网络及有向加权网络

2.1 复杂网络

复杂网络借助图论和统计物理方法来捕捉并描述系统的整体行为。在复杂网络描述中,节点表示系统的组成元素,边表示系统诸多组成元素之间的关系,抽象为二元组 $G=(V,E)$,其中节点总数 $N=|V|$,边总数 $M=|E|$ 。若点对 (i,j) 与 (j,i) 对应同一条边,则该网络为无向网络(undirected network),否则为有向网络(directed network);若每条边都有各自相应的权值即为加权网络(weighted network),否则为无权网络(unweighted network)^[1-3]。

目前国内外对无向无权复杂网络研究较多,主要围绕节点的度(degree)及度分布(degree distribution)、平均路径长度(average path length)、簇系数(clustering coefficient)、度相关性(degree correlation)、介数(betweenness)等指标进行研究。此外对复杂网络社团(community structure)研究也较多。

2.2 有向加权网络

在有向加权复杂网络中,边权 W_{ij} 是指从节点 V_i 到节点 V_j 连接的权重, S_i 表示节点 V_i 的强度(strength)。节点 V_i 的强度指与此节点连接的边的权重之和。节点的人强度为从其他节点指向该节点的所有的边的权重之和,节点的出强度为从该节点指向其他节点的所有的边的权重之和。节点的人强度和节点的出强度具体定义如下^[4-6]:

定义 1 节点的人强度

$$S_i^{\text{in}} = \sum_{j \in V(i)} W_{ji} \quad (1)$$

定义 2 节点的出强度

$$S_i^{\text{out}} = \sum_{j \in V(i)} W_{ij} \quad (2)$$

其中, $V(i)$ 为节点 V_i 的近邻集合。根据定义 1

和定义 2,节点的总强度为节点的人强度和节点的出强度之和。

定义 3 节点的总强度

$$S_i = S_i^{\text{in}} + S_i^{\text{out}} \quad (3)$$

节点的强度分布用 $P(K)$ 描述,表示一个随机选定的节点的强度恰好为 k 的概率,即强度为 k 的概率 $P(K)$ 随节点强度值 k 的变化。同理,用 $P_{\text{in}}(K)$ 和 $P_{\text{out}}(K)$ 分别描述节点的人强度分布和节点的出强度分布。统计表明,节点的强度分布、节点的人强度分布和节点的出强度分布主要有两种形式:一是指数分布(Poisson 泊松分布),二是幂律分布。

有向加权复杂网络中,两节点间的最短路径是指在连接两节点的连通路程中,权重之和最小的一条或几条路径。如果两节点不存在连通路程,则这两点间的路径长度为 ∞ 。由于在有向加权复杂网络中,可能存在两节点无连通路程的问题(重点是由于节点连接的方向问题),所以,我们认为在有向加权复杂网络中研究其平均最短路径不太可能。

节点的簇系数(clustering coefficient)用于衡量复杂网络集团化程度,即复杂网络中节点之间的有向连通程度,或网络节点的局部传递程度。节点的近邻之间联系越紧密,该节点的簇系数越高。节点的介数(betweenness)用于度量复杂网络中所有最短路径中通过该节点的路径数量。在有向加权复杂网络中,通过某节点的最短路径越多,该节点的重要程度越高。高介数的节点是网络的主要节点,表征了节点在整个网络中的全局作用。节点的介数定义如下:

$$\text{定义 4} \quad B_u = \sum_{i \neq j} \frac{\delta(i,u,j)}{\delta(i,j)} \quad (4)$$

其中, $\delta(i,j)$ 表示节点 i 和 j 之间总的最短路径的个数, $\delta(i,u,j)$ 表示节点 i 和 j 之间经过节点 u 的最短路径的个数。

3 软件系统复杂网络特性

最近几年,研究者借助复杂网络对大量软件系统的类图进行研究,发现软件系统的结构及内部通信过程并不是随机和无序的,大多具有“小世界”(small-world)和“无标度”(scale-free)等复杂网络特性,即具有较大的簇系数和较高的介数,以及节点的连接呈指数分布或幂律分布^[4-5]。

通过逆向工程对部分软件系统的组成元素及整体结构进行研究,发现早期自由设计的软件规模较小,没有遵循软件工程开发思想,受程序设计者的个人影响较大,软件呈现随机网络特征;而在开发中遵

循了软件工程的层次化、高内聚、低耦合等开发设计思想的系统,控制结构趋于严格,具有规则网络的部分特性,并具有复杂网络的许多特征(如“小世界”、“无标度”等)。研究团队还发现,只要能保持系统诸组成元素(如构件、类)之间交互的基本性质不变,那么系统的基本特性几乎不会改变。

把一个实际的软件系统抽象为有向加权网络的过程所面临的一个重要问题是边权的赋予方式。目前,国内外边权赋予方式主要分为三类:第一类是基于“功能驱动”原理的赋权法;第二类是基于“差异驱动”原理的赋权法;第三类是综合集成赋权法^[7-8]。权重又可分为相异权(dissimilarity weight)和相似权(similarity weight)两种。相异权的权值越大表示两节点之间的距离越大,关系越疏远;相似权的权值越大表示两节点间的关系越亲密,距离越小。

由于对软件系统整体结构和整体运行过程的把握及了解程度,直接影响软件系统复杂性的准确度量,而 UML 的用例图即是从系统整体角度及整体运行过程出发,表达若干执行者(actor)以及这些执行者与系统提供的用例之间的连接执行关系。鉴于此,论文提出一种基于 UML 用例图的软件系统复杂网络特性度量方法,具体如下:

步骤 1:分析系统用例图,估计每个用例的执行概率;

步骤 2:将用例图转化为有向加权复杂网络,用例的执行概率即为有向加权网络的边权。执行概率转化为相应边权的算法为:边权=父用例执行概率×本用例执行概率;

步骤 3:利用定义 1~定义 4 进行转化后网络的复杂网络特性度量,度量值即为该软件系统的复杂网络特性值。

文中主要围绕节点的强度、入强度、出强度、节点的强度分布、节点的入强度分布、节点的出强度分布、节点的介数等几个方面进行度量研究。通过度量,节点的入强度越大,说明与该节点对应的模块(或构件)被较多模块所调用,或较多模块依赖于该模块;节点的出强度越大,说明与该节点对应的模块要依赖于较多模块(隐含了调用结果值的返回)。同理,与节点对应的模块的介数高低,可反映该模块在整个软件系统中的重要程度。

4 实例分析

随着 3S 技术的发展,在国家科技部的大力支持下,结合新疆生产建设兵团(以下简称“兵团”)实际,

兵团开发了兵团区域空间信息服务系统。该系统包括农业、水利、土地、生态环境、公共安全等五个子系统,分别服务于兵团、各职能部门、师(局)、农牧团场、连队,实现各个部门对信息的及时了解及部门之间的信息共享。在兵团区域空间信息服务系统开发中,必须为农业、水利、土地、生态环境、公共安全等五个子系统提供一个空间信息管理与服务平台,实现各个子系统的集成及各类信息的共享。由于本系统非常庞大,本文只选取该管理与服务平台进行实证研究。

空间信息管理与服务平台包括:数据转换、图形矢量数据编辑、显示控制、查询统计、空间分析、三维模拟与仿真、元数据管理、系统管理等八个功能模块,这八个功能模块为整个系统的运行提供了支持。

结合系统用例图并通过实际分析,空间信息管理与服务平台 8 个功能模块的执行概率分别为 5%、5%、15%、30%、16%、9%、15%、5%。该平台功能较多,本文只对 8 个功能模块中的元数据管理模块进行进一步研究。经分析,元数据管理模块中元数据模式管理、元数据建库、元数据录入编辑的运行概率又分别为 20%、20%、60%。依此类推,可得出各个功能的运行概率,如图 1 所示。

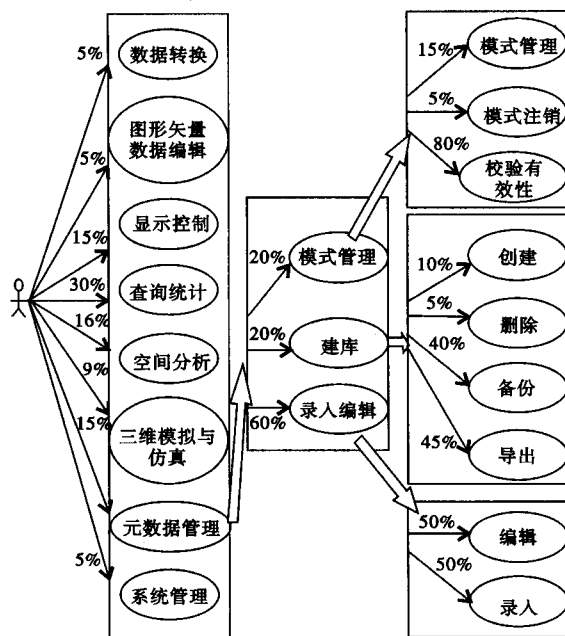


图 1 空间信息管理与服务平台用例图

将图 1 的用例图及每个用例的执行概率转化为有向加权复杂网络图如图 2 所示。转化后的网络用二元组 $G=(V,E)$ 表示,其中节点集合 $V=\{V_1,V_2,\dots,V_{20}\}$,有向边集合 $E=\{E_1,E_2,\dots,E_{20}\}$ 。边权权

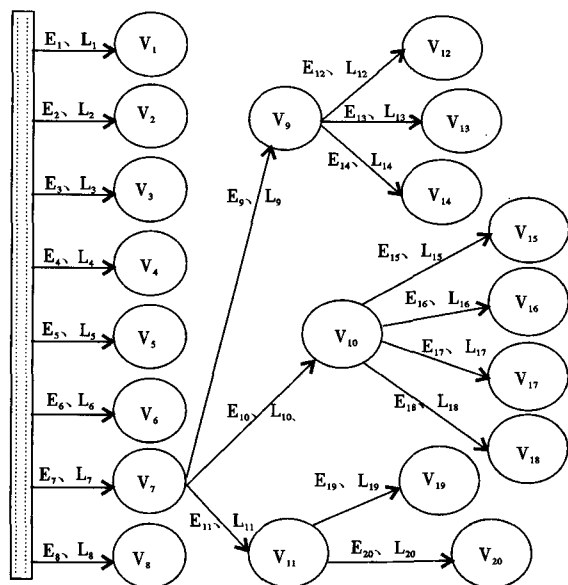


图2 系统复杂网络图

重集合 $L = \{L_1, L_2, \dots, L_{20}\}$. 在转换过程中, 执行概率转化为相应边权的算法为: 边权 = 父用例执行概率 $\times$ 本用例执行概率. 如, $L_9 =$ 元数据管理用例执行概率 $\times$ 模式管理用例执行概率, $L_{15} =$ 元数据管理用例执行概率 $\times$ 建库用例执行概率 $\times$ 创建用例执行概率. 依此类推, 得到每条边的边权如表 1 所示.

根据图 2 和表 1 所知, 20 个节点的入强度分别为: $\bar{S}_1 = 0.05, \bar{S}_2 = 0.05, \bar{S}_3 = 0.15, \bar{S}_4 = 0.30, \bar{S}_5 = 0.16, \bar{S}_6 = 0.09, \bar{S}_7 = 0.15, \bar{S}_8 = 0.05, \bar{S}_9 = 0.030, \bar{S}_{10} = 0.030, \bar{S}_{11} = 0.090, \bar{S}_{12} = 0.0045, \bar{S}_{13} = 0.0015, \bar{S}_{14} = 0.0240, \bar{S}_{15} = 0.0030, \bar{S}_{16} = 0.0015, \bar{S}_{17} = 0.0120, \bar{S}_{18} = 0.0135, \bar{S}_{19} = 0.0450, \bar{S}_{20} = 0.0450$.

表1 复杂网络边权表

边	边权	边	边权
L_1	0.05	L_{11}	$A \times 0.6 = 0.090$
L_2	0.05	L_{12}	$B \times 0.15 = 0.0045$
L_3	0.15	L_{13}	$B \times 0.05 = 0.0015$
L_4	0.30	L_{14}	$B \times 0.80 = 0.0240$
L_5	0.16	L_{15}	$B \times 0.10 = 0.0030$
L_6	0.09	L_{16}	$B \times 0.05 = 0.0015$
L_7	0.15	L_{17}	$B \times 0.40 = 0.0120$
L_8	0.05	L_{18}	$B \times 0.45 = 0.0135$
L_9	$A \times 0.2 = 0.030$	L_{19}	$C \times 0.50 = 0.0450$
L_{10}	$A \times 0.2 = 0.030$	L_{20}	$C \times 0.50 = 0.0450$

注: $A = 0.15, B = A \times 0.2, C = A \times 0.6$

同理, 可计算出每个节点的出强度和总强度值. 研究发现, 本网络同一层次节点 (如 $V_1 \sim V_8$, 或 $V_{12} \sim V_{20}$) 的入强度分布 $P_{in}(K)$ 、出强度分布 $P_{out}(K)$ 和总强度分布 $P(K)$ 服从 Poisson 分布.

由于本系统比较特殊, 两个节点的最短路径的个数与这两点经过中间节点的最短路径的个数相同, 所以, 节点的介数值恒为 1.

同理, 可借助本方法度量本系统其他子系统及整个系统的复杂网络特性.

5 结束语

文中基于 UML 的用例图进行系统整体运行过程及整体角度分析, 将用例图和复杂网络理论相结合进行软件系统复杂网络特性度量, 并主要围绕节点的入强度、出强度、节点的强度分布、节点的入强度分布、节点的出强度分布、节点的介数等几个方面进行度量研究. 最后, 通过兵团空间信息系统的管理与服务平台进行实证研究. 实践表明, 此方法能较好地对软件系统的复杂网络特性进行度量, 能有效提高所开发系统的质量及可靠性.

参考文献:

- [1] W Li W, Hong Wang, Na Qiu. Local-word evolving model of directed and weighted network[J]. Computer Engineering and Application, 2010(4-6): 94-97.
- [2] Wang G-Y, Jian Zhou, Yang Xie. Directed weighted network model based on BBV[J]. Computer Engineering, 2010, 36(12): 141-143.
- [3] Newman M E J. Modularity and community structure in networks [J]. Natl. Acad. Sci. USA, 2006(103): 8577-8582.
- [4] Concas G M. Power-laws in a large object-oriented software system [J]. IEEE Trans. on Soft. Eng. 2007, 33(10): 687-708.
- [5] Zhang C C, Qi X G. Analysis of wireless sensor network characteristics measurement based of complex network theory[J]. Computer Science, 2011, 37(2): 44-46.
- [6] Bing Li, Ma Y T, Jing Liu. Advances in the studies on complex networks of software systems[J]. 2008(36): 805-814.
- [7] 秦怀斌, 梁斌, 邵明文. 软件体系结构复杂网络社团特性研究[J]. 微电子学与计算机, 2012, 29(4): 6-9.
- [8] 文静, 顾勋梅. 一种度量软件需求的功能点方法[J]. 微电子学与计算机, 2011, 28(10): 75-78.

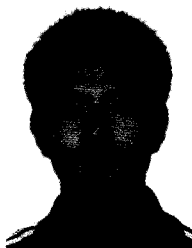
(下转第 80 页)

的顶点处理器,采用多线程结构,可以有效地减少发生数据读写冲突时的流水线停顿周期数,提高了处理效率.在相对较少硬件开销的前提下,性能有了较大地提升.

参考文献:

- [1] Khronos Group. OpenGL ES common profile specification[M]. USA: Khronos Group, 2008.
- [2] Microsoft. Direct 3D[EB/OL]. [2011-05-17]. [http://msdn.microsoft.com/en-us/library/windows/desktop/hh309466\(v=VS.85\).aspx](http://msdn.microsoft.com/en-us/library/windows/desktop/hh309466(v=VS.85).aspx).
- [3] Hill F S, Kelley S M. 计算机图形学(OpenGL 版)[M]. 3 版. 胡事民,译. 北京:清华大学出版社,2009.
- [4] Sohn Ju-Ho, Woo Jeong-Ho, Lee Min-Wuk, et al. A 155-mW 50-mvertices/s graphics processor with fixed-point programmable vertex shader for mobile applications[J]. IEEE Journal of Solid-State Circuits, 2006, 41(5): 1081-1091.
- [5] 杨毅,郭立,史鸿声,等. 一个面向移动设备的可编程顶点处理器的设计[J]. 中国科学技术大学学报, 2010, 40(2): 140-145.
- [6] Arakawa Fumio, Yoshinaga Takeshi, Hayashi Tomochi, et al. An embedded processor core for consumer appliances with 2.8GFLOPS and 36M polygons/s FPU[C]//Digest of Technical Papers. USA: IEEE ISSCC, 2004: 334-343.
- [7] Kim Donghyun, Chung Kyusik, Yu Chang-Hyo, et al. An SoC with 1.3 gtexels/s 3-D graphics full pipeline for consumer applications[J]. IEEE Journal of Solid-State Circuits, 2006, 41(1): 71-84.
- [8] Yu Chang-Hyo, Chung Kyusik, Kim Donghyun, et al. An energy-efficient mobile vertex processor with multithread expanded VLIW architecture and vertex caches[J]. IEEE Journal of Solid-State Circuits, 2007, 42(10): 2257-2269.
- [9] Lindholm Erik, Kilgard Mark J, Moreton Henry. A user-programmable vertex engine[C]//Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques. Los Angeles, CA, USA: ACM SIGGRAPH'01, 2001: 149-158.
- [10] 韩亮,陈杰,史鸿声,等. 嵌入式可重构 DSP 处理器的指令译码器设计[J]. 微电子学与计算机, 2004, 21(7): 91-94.
- [11] 敖天勇,陈杰,刘建,等. 基于 Costar II 的异构多核 DSP 设计与实现[J]. 微电子学与计算机, 2007, 21(3): 59-62.

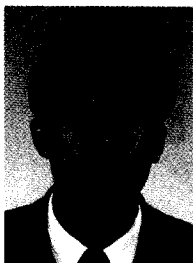
作者简介:



申智波 男,(1987—),硕士研究生.研究方向为集成电路设计.

(上接第 75 页)

作者简介:



秦怀斌 男,(1980—),博士研究生,副教授.研究方向为软件工程、复杂网络.



郭理 男,(1968—),硕士,副教授.研究方向为软件工程、人工智能.