软件逆向工程技术研究

严 秀[1,2], 李龙澍[1,2]
(1. 安徽大学 计算机科学与技术学院, 安徽 合肥 230039;
2. 安徽大学 计算智能与信号处理教育部重点实验室, 安徽 合肥 230039)

摘 要: 随着计算机技术的发展, 软件系统的规模和复杂度日益增长, 软件维护在整个系统开发过程中愈发重要, 越来越多的遗产系统需要维护和改善, 逆向工程已经成为软件维护的关键技术之一。介绍了逆向工程的基本概念, 综述了主要步骤和分析方法, 最后通过分析逆向工程在国内外的研究现状, 指出了存在的一些问题, 并给出了未来的发展趋势。

关键词: 软件维护; 遗传系统; 逆向工程
中图分类号: TP311.5 文献标识码: A 文章编号: 1673-629X(2009)04-0020-05

Research of Technology in Software Reverse Engineering

YAN Xiu[1,2], LI Long-shu[1,2]
(1. School of Computer Science and Technology, Anhui University, Hefei 230039, China;
2. Ministry of Edu. Key Lab. of Intelligent Computing & Signal Processing at Anhui Univ., Hefei 230039, China)

Abstract: With the development of computer technology, the size and complication of the software system increase progressively, which brings software maintenance become more important. More and more legacy applications must be maintained and improved. Reverse engineering has become one of the crucial techniques in software maintenance. Introduces the basic concept of the reverse engineering firstly, then summarizes its main process and analyzing method. Finally by analyzing the research status of reverse engineering, a discussion of the drawbacks of reverse engineering and a comment on the future of reverse engineering is presented.

Key words: software maintenance; legacy system; reverse engineering

0 引 言

随着计算机技术的迅速发展, 计算机技术应用的领域也逐渐扩大, 人们希望计算机这一智能体能够解决各个领域的更多、更复杂的问题, 从而也对计算机软件产品的功能、质量、开发成本和时间提出了越来越多的要求, 软件技术受到了前所未有的挑战。传统的软件工程主要关注新软件的分析与设计, 但随着软件系统的规模和复杂度日益增长, 软件的生命周期越来越长, 软件开发的很大一部分工作集中于维护和改造现有的软件系统, 而这些现有系统的需求、设计决策、业务规则、历史数据等统称为遗产系统(LS[1], Legacy System), LS 是一种巨大的、长期的投资, 因为如何充分利用这些有用的资产对新系统的开发显得尤其重要。另一方面, 随着 Internet 技术的普及, Web 用户增多, 很多软件厂商需要将系统移植到 Web 上, 进一步加剧了对软件维护的需求。实践研究表明, 软件资源预算的 50%~80% 消耗在对现有系统的维护上[2], 而软件维护者理解程序源代码的时间要占整个软件维护的 47%~62%[3]。软件维护已经成为软件工程面临的重要课题之一, 而正确和全面地理解软件系统是对软件进行维护的前提, 软件逆向工程应运而生, 成为软件工程领域的一个新兴分支, 其目标就是开发帮助人们理解已有软件系统的方法、工具, 为软件系统的维护和演化提供支持。

文中主要介绍逆向工程的基本概念、主要步骤、分析方法、研究现状、存在的问题以及发展方向。

收稿日期: 2008-08-06
基金项目: 国家自然科学基金项目(60273043); 安徽省高校拔尖人才基金项目(05025102); 安徽省自然科学基金项目(050420204); 安徽省教育厅自然科学研究基金项目(2006KJ098B)
作者简介: 严 秀(1985-), 女, 硕士研究生, 研究领域为智能软件; 李龙澍, 教授, 博士生导师, 研究领域为智能软件和知识工程。

1 逆向工程的基本概念

“逆向工程”[4]这个名词最早出现在对硬件产品的分析中, 人们分析硬件产品以便改进自己的产品, M. G. Rekoffjr[5] 将逆向工程定义为: 对一个复杂的硬件系统实施有条理的检查, 以开发出关于这个系统的一组

规范说明的过程。在把这个概念应用到软件系统过程中, 研究人员发现利用其中的许多方法可以获得对系统以及系统结构的理解。然而, 对一个硬件系统实施逆向工程, 一般是为了得到这个系统的复制品, 对一个软件系统实施逆向工程, 一般是为了获得对这个系统在设计层次上的理解, 以便于系统的维护、巩固、移植、改进。

软件逆向工程的基本原理是抽取软件系统的主要部分而隐藏细节, 然后使用抽取出的实体在高层上描述软件系统。在软件工程领域, 迄今为止没有统一的逆向工程定义, 较为通用的是 Elliot Chikofsky 和 Cross[6] 1990 年在文献[7] 中定义的逆向工程的相关术语。软件工程通常被认为是开发一个新的系统, 尽管软件工程也包括逆向工程和再工程, 为了避免对软件工程含义的误解, 引进了正向工程的概念。

(1) 正向工程(Forward Engineering): 从系统的高层抽象和逻辑上独立于实现的设计到系统的物理设计的传统过程, 具体地说是从用户的需求到高层设计, 再到底层设计, 最后到实现的过程。

(2) 逆向工程[8] (Reverse Engineering): 对系统进行分析, 以确定系统的组件和组件之间的相互作用, 以其他形式表示系统, 或在较高的抽象层次上表示系统的过程。值得说明的是, 在对一个系统实施逆向工程时, 并不改变这个系统本身, 也不包括在此系统上构建新的系统的过程。

(3) 重构[9](Restructuring): 保持系统外部行为(功能和语义) 的前提下, 在统一抽象层次上改变表示形式。

(4) 再工程[10] (Reengineering): 通过逆向工程、重构和正向工程对现有系统进行审查和改造, 将其重组为一种新形式。

(5) 设计恢复(Design Recovery): 结合目标系统、领域知识和外部消息认定更高层次的抽象。

其中, 再工程、设计恢复不改变系统, 重构改变了系统, 但不改变其功能, 再工程涉及到正向工程与逆向工程的联合使用, 逆向工程解决程序的理解问题, 正向工程检验哪些功能需要增加、保留和删除, 再工程改变了系统的功能和方向, 是最根本和最有深远影响的扩展。

图 1 显示这些概念之间的关系[11]。

2　主要步骤和分析方法

由逆向工程定义[12] 可知: 软件逆向工程的任务包括分析系统、抽象系统和展现系统, 从而实现协助用户理解系统的目的。

分析系统是指分析系统的结构及运行过程, 但不管目标系统面向何种应用领域, 分析系统不外乎是分析系统的静态信息和动态信息。目标系统面对不同的应用领域, 要实现抽象目标系统的任务, 需要领域知识和专家的经验。展现系统最好的方式是使系统可视化。

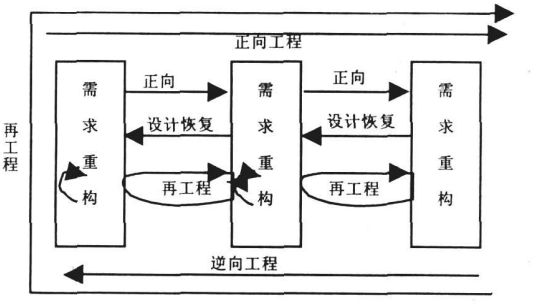

图 1　关系图

现有的逆向工程分析方法[13] 主要有以下 4 种:

(1) 词法分析和语法分析。

该方法主要是对程序源码进行分析, 得到程序信息的多种有用表示, 其中最常用的就是交叉引用列表。通过语法分析可以得到两类表示: 分析树(parse tree)、抽象语法树 AST(abstract syntax tree), 其中 AST 是更复杂的程序分析工具基础, 包含了和程序的实际内容相关的细节。

(2) 图形化方法。

图形化方法包括控制流分析、数据流分析以及程序依赖图。控制流分析是在确定程序语法结构之后进行。数据流分析关注于解决程序中从定义到使用的过程的相关的问题, 比控制流分析要复杂得多。程序依赖图是数据流分析的进一步改进, 比数据流分析更复杂。在程序依赖图中, 控制流和数据流依赖放在一起处理, 程序依赖图还具有这样的结构特性: 一个程序依赖图描述了一个控制依赖的区域。

(3) 程序切片。

切片技术来源于数据流分析方法, 已经成为很多程序理解工具的基础。一个程序切片是由程序中的一些语句和判定表达式组成的集合。这些语句和判定表达式可能会影响在程序的某个位置上所定义或使用的变量的值。利用切片技术可以将关注点确定在一个较小范围而不是关注整个程序。

(4) 动态分析。

静态分析是对程序源码进行分析。动态分析则是在程序运行时进行分析, 基本方法是对程序进行植入。植入是在一种在全局范围内更改源代码以添加额外操作的过程。这种方法的基本原理是: 利用代码的结构信息, 依据固定的规则, 将软件触发器添加到代码中。

3 研究现状

软件逆向工程的研究已经有 10 多年的历史了。在国外,软件逆向工程是作为对软件维护的一部分出现的,主要是通过逆向工程理解程序,对系统进行维护、迁移和进化遗产系统。

目前,逆向工程技术的重要性已经引起重视,得到了国内外学术界和商业界的广泛认同。在学术界,面向逆向工程领域的国际会议 WCRE[14] (the Working Conference on Reverse Engineering)、IWPC (the International Workshop on Program Comprehension) 和 PASTE (the Workshop on Program Analysis for Software Tools and Engineering) 每年举行一次。卡内基梅隆大学软件工程研究所成立了专门的再工程中心,致力于逆向工程的研究。IBM 研究中心(IBM Research) 设立了“软件工程中关注点的多维分解”研究项目,研究工作已经进行了多年。

逆向工程技术发展至今,已经研制开发出许多工具,下面介绍一些典型的国内外逆向工程工具。

(1) Rational Rose[15] / Rose RealTime/ Rose/ Architect。

Rose/ Rose RealTime(Rose I) 提供的逆向工程工具,可以从多种程序设计语言源程序中自动产生静态设计模型,但目前只能逆向产生类图。Rose/ Architect 是 USC(University of Southern California) 与 Rational 合作开发的一种可视化工具,用于对 UML 类图中的实体进行基于规则的等价合并,以突出地呈现系统的软件体系结构成份。

(2) Rigi[6]。

Rigi 是一个可扩展、可裁剪的逆向工程环境。用半自动的工具从软件表示中提取数据信息,将信息存入低层库中,系统被抽象为子系统的分层结构。主要由 3 部分构成: 支持 C/ C+ + 、COBOL 等语言的程序静态信息解析器用于存储从源代码提取的信息的程序静态信息库以及展示和操纵程序静态信息的交互式窗口图形编辑器。Rigi 可以与一种支持面向对象动态建模的环境 SCED 协同使用,分别得出目标系统源程序的静态信息和动态信息。

(3) Microsoft Visio 2002。

有三个不同版本: 标准版提供图表解决方案,帮助业务专业人员共享他们日常处理的信息并加以可视化;专业版增加了新的功能,帮助技术专业人员对现有创意、信息和系统加以可视化,并建立新创意、信息和系统的原型;企业版带有企业网络工具附件,IT 专业人员可以获得高级网络图表和文档能力。

(4) Sniff+。

Sniff+ 不是一个单纯的逆向工程工具,而是一个开放的、可扩展、伸缩的 C/ C+ + 程序设计环境,具有逆向工程能力。Sniff+ 提供了一个有效而轻便的环境,用户界面友好。Sniff+ 适用于不能完全解析的半成品软件系统,在产生可打印的视图和浏览半成品系统等方面也具有特色。

(5) Imagix4D。

一个 C/ C+ + 程序的静态理解[16] 工具,它可以提供多层次的视图,以表示从高层设计到实现的细节、类和函数的依赖关系等。Imagix4D 以三维图形的形式展示关键信息,使得用户的注意力集中在回答特定问题的视图上。Imagix4D 提供了产生大量视图的能力以及从源代码生成文档的能力。它提供的交互式询问功能对迅速理解程序很有帮助。

(6) MORALE/ ISVis。

MORALE(Mission Oriented Architectural Legacy Evolution) 是 DARPA 资助的一个 EDCS(Evolutionary Design of Complex Software, 复杂软件的进化式设计) 类课题。课题中与逆向工程有直接关系的是 ISVis (Interaction Scenario Visualize) 工具。这种工具的功能包括: 读入源程序文件,解析后产生静态信息文件;利用源代码、静态信息文件以及用户提供的相关的信息,产生获取动态信息所需的代码;执行插入上述代码后的目标系统,产生动态事件跟踪信息,并自动转变成剧情;用户通过所产生的剧情视图进行交互式分析。

(7) August- II。

August- II 是一个数据逆向工程工具。该工具以 COBOL 记录格式或 DB2 数据定义等各种各样的数据资源为输入,产生概念数据模型,使用户能够理解当前的环境并转向新的数据技术。其输出能用作许多不同软件包如 CASE 工具和商业数据库管理系统的输入。

(8) Refine/ C。

一种可扩展的、交互式反转 C 程序的工作平台,它提供了应用程序设计接口,以支持用户建立具有自己风格的源程序分析工具。在许多国内外文献中对 Refine/ C 程序解析器评价很高,并认为它具有良好的可扩展性。

国内的逆向工程工具主要有青鸟程序理解系统 JBPAS (Jade Bird Program Analysis System), 是青鸟 II 型系统的组成部分。JBPAS 是由一个 C+ + 分析器前端和一组分析工具组成的程序理解系统。该系统针对 C+ + 语言,采用增量分析技术对程序源代码进行静态分析;用 EER(Enhanced Entity Relationship) 为 C+ + 程序建立概念模型并抽取程序信息,将信息保存在数据库中;按照不同的用户需求组织程序信息,辅助用

户理解 C+ + 程序; 逆向生成源程序的 OOD(Object-Oriented Design) 文档和 Rose 文档。该系统中的面向对象测试支持工具(Object- Oriented Testing Supporter) 能够利用插装技术跟踪程序的运行, 以辅助测试用例的生成。

4　存在问题

尽管经过这么多年的发展, 逆向工程研究取得了进步性的研究成果[17], 但在应用、理论方面仍然不够成熟, 有待于在以下几方面进行进一步研究:

(1) 符合经济实用的需求。

逆向工程的目标在于为系统维护和系统演化中的系统理解提供支持, 其在实际使用中的经济影响是至关重要的。因而, 逆向工程系统必须以实用性为首要目的, 要能产生实际的经济效益。逆向工程过程相当复杂, 开发全自动的逆向工程工具解决真正的问题在短期内是不可能的。所以, 逆向工程工具的研究应注重于实用的半自动工具的开发, 并做大量的案例研究。利用逆向工程进行开发初期就应对经济效益作评估, 不要一味追求这一方法而导致经济损挫。

(2) 缺乏统一的逆向工程的概念、标准术语, 缺乏统一的逆向工程机制的分类框架, 缺乏对现有逆向工程的广泛使用, 缺乏对现有工具和理论进行有效评估的标准及工具, 这些都导致了研究人员在交流上的困难, 不利于工具的研制, 也不利于逆向工程技术的应用、推广和提高。

(3) 避免使用虚构的数据。

由于其不成熟, 很多逆向工程研究都针对于虚构问题。但是, 以虚构问题为目标限制了逆向工程系统的实用性。为了使逆向工程的研究真正走向成熟, 必须从虚构问题过渡到有经济影响的实际问题。为此, 逆向工程研究应当遵循以下几个原则: a. 用实际程序作为例子; b. 用系统作为例子; c. 利用多种信息源。

(4) 黑盒理解。

传统的逆向工程辅助程序理解的方式可视为"白盒"理解, 主要依赖于分析源代码程序抽取程序结构和控制流信息。由分布式对象技术、构件技术和构架技术发展而导致的基于构件的软件开发对程序理解提出了新的挑战。逆向工程的黑盒理解不仅可用于传统的系统维护, 而且可用于新构件的验证、适应性修改和组装。很多情况下, 系统的成功演化需要结合使用白盒理解和黑盒理解。

5　发展方向

逆向工程研究作为软件工程中一个正在兴起的研究领域, 理论和应用的研究都处于探索阶段, 笔者认为应该在以下几个方面作进一步的研究:

(1) 与具体领域相结合。对一个特定系统进行逆向工程的根本目的是理解这个系统的结构和行为, 而这个系统的结构和实现过程必然受它所处理的问题的影响, 所以要很好地理解这个系统, 一些领域知识是必不可少的。在逆向工程工具中可以用设立领域知识库, 增加人机交互等手段解决这个问题。

(2) 提高逆向工程过程的可重复性[18], 使逆向工程过程可定义、可管理及可优化, 要注重于提高逆向工程过程的自动化程度, 开发实用的半自动化的工具。尽量使用工具自动实现逆向工程过程的各部分功能, 减少用户的负担。

(3) 增加学习功能[19]。对目标系统实施逆向工程得到的结果不一定完全满足用户的要求, 需要将满足用户要求的部分结果保留, 对不满足要求的部分进一步实施逆向工程, 经过多次迭代得到用户最终满意的结果, 这就需要在逆向工程中增加学习功能。

(4) 随着 Web 应用的普及, 需要对 Web 应用程序的逆向工程方法作进一步的研究。近几年来, 开放源代码逐渐成为一种趋势, 为了达到软件架构[20] 和设计模式的复用, 从得到的源代码中获取软件设计模式和架构模式也将成为广泛的需求, 这就需要逆向工程的支持, 从这里可以看到逆向工程广泛的应用前景。

(5) 改善逆向工程工具的性能, 使工具作为一项功能附加到成熟的开发环境中。工具还必须容易使用, 能够提供方便、有效的功能, 吸引更多的用户。逆向工程工具只有在使用过程中才能真正发现问题, 得到更好的发展。

6　结束语

综上所述, 逆向工程作为一个新兴的领域, 在软件维护[21] 中有着重要的作用, 充分利用逆向工程技术就可以对现有系统进行改造, 减少开发强度, 提高软件开发效率, 降低项目开发的经济成本, 提高经济效益, 并在一定程度上保证软件开发和利用的延续性。

有关逆向工程的信息可以从以下几方面获得: 有关软件维护和再工程的欧洲会议(CSRE) , 逆向工程工作会议[22] (WCRE) , 逆向工程和再工程 IEEE TCSE 委员会(这个委员会致力于现存软件系统的检查技术和软件再工程的方法) , 再工程论坛(Reengineering Forum) 及再工程工作组等。

参考文献:

[1]　宋海鸿, 陈　平. 逆向工程在软件开发中的作用和应用现

状[J]. 电子科技, 2002(1): 28- 30.
[2] Boehm B W. Software engineering economics[M]. [s. l.]: Prentice Hall PTR, 1981.
[3] Fjeldstad R K, Hamlen W T. Application program maintenance study: Report to our respondents[C]//Proceedings GUIDE 48. Philadelphia: [s. n.], 1983.
[4] 王玉英, 陈 平, 方海燕, 等. 软件逆向工程的研究与发展[J]. 西安工程科技学院院报, 2006(6): 374- 375.
[5] Rekoffjr M G. On reverse engineering[J]. IEEE Trans systems, man, and cybernetics, 1985, 18(2): 244- 252.
[6] Chikofsky E J, Cross J H. Reverse engineering and design recovery: A taxonomy[J]. IEEE Software, 1990, 7(1): 13- 17.
[7] Pinzger M. Harald gall: Pattern- supported architecture recovery[C]//IWPC. Paris: [s. n.], 2002: 53- 64.
[8] 周立萍, 陈 平. 逆向工程发展现状研究[J]. 计算机工程与设计, 2004(10): 1658- 1660.
[9] 袁望洪, 陈向葵, 谢 涛, 等. 逆向工程的研究与发展[J]. 计算机科学, 1999, 26(5): 71- 77.
[10] 郭 颖, 钱 渊. 逆向工程的应用研究和发展[J]. 信息与电子工程, 2004(6): 157- 158.
[11] 李青山. 面向对象软件的动态模型设计恢复与体系结构抽象[D]. 西安: 西安电子科技大学, 2003.
[12] 李伟华, 李 由. 实时软件逆向工程技术研究[J]. 西北工业大学学报, 2004(3): 392- 394.
[13] Muller Hausi A, Smith Dennis B. Reverse engineer: A roadmap[C]//Proceedings of Future of Software Engineering. New York: [s. n.], 2000.
[14] Demeyers, Ducasses, Nierstasao. Object- Oriented Software Reengineering[EB/OL]. [2004- 04- 05]. http://www.iam.unibe.ch/scg/Archive/Lectures/OOSR- W99.pdf.
[15] Di Lucca G A, Di Penta M, Antoniol G. An approach for reverse engineering of web- based applications[C]//Proceedings of WCRE' 01. Washington: IEEE Computer Society Press, 2001: 231- 240.
[16] Davis, Kathi. Hogshead, august- II: A tool for step- by- step data model reverse engineering[C]//Proceedings of the Second Working Conference on Reverse Engineering. [s. l.]: [s. n.], 1995: 146- 155.
[17] 张志猛. 面向对象软件的逆向工程[J]. 计算机研究与发展, 2003, 40(7): 1062- 1068.
[18] Biggerstaff T J. Design recovery for maintenance and reuse[J]. IEEE Computer, 1989, 22(7): 36- 49.
[19] Bellay B, Gall H. A comparison of four reverse engineering tools[C]//Proceedings of WCRE' 97. Amsterdam: [s. n.], 1997: 2- 12.
[20] Bisbal J. Legacy Information Systems: Issues and Directions[J]. IEEE Software, 1999, 10(9): 103- 110.
[21] 郭 耀, 袁望洪, 陈向葵, 等. 再工程——概念及框架[J]. 计算机科学, 1999(5): 78- 80.
[22] Su Yang, Li Fan, Hu Sheng- ming, et al. Aspect- oriented software reverse engineering[J]. Journal of Shanghai University (English Edition), 2006, 10(5): 402- 408.

(上接第 19 页)

非常小, 有利于 FPGA 利用剩下的资源更好地完成轨迹关联算法等的设计实现, 大大节省了硬件资源。

实验中, 输入的二值图像尺寸为 256 × 256, 帧频为 100 帧/秒, 经过能量积累及门限分割后的二值图像虚警概率为 0.3%。目标关联检测选取窗口大小为 32 × 32, 每个窗口内平均存在 3 个候选点, 经过目标关联检测, 图像虚警概率降为 0.1%, 剔除了大部分噪声点, 达到了预期的目的, 大大减少了参与轨迹关联的候选目标点数目, 提高了轨迹关联的精度和速度, 为轨迹关联算法的硬件实现打下了良好的基础。

4 结束语

为便于轨迹关联算法的硬件实现, 文中提出目标关联检测算法, 用于对图像进行轨迹关联前的预处理, 并给出了该算法的 FPGA 实现以及实验结果。

文中的设计结构简单、流程清晰, 提出了适合硬件实现的流水线结构, 提高了数据处理速度。实验结果表明设计运行速度快、占用资源少, 具有很强的实用性。

实验采用连续的 256* 256 的图像序列, 目标关联检测算法对能量累加及门限分割后图像中的噪声点予以进一步剔除, 使图像的虚警概率降低了 70%, 大大减少了轨迹关联部分的运算量, 在目标关联检测后采用轨迹关联可以有效地检测出目标, 检测概率 ≥98%。

参考文献:
[1] Zhang Bing, Lu Huangzhang. The predicting and matching detection algorithm of moving point target in image sequences[C]//IEEE International Conference on Neural Networks & Signal Processing. Nanjing, China: [s. n.], 2003.
[2] 孙德宝, 周卫祥. 红外图像序列运动小目标检测的预处理算法研究[J]. 红外与激光工程, 2000, 29(2): 12- 14.
[3] 黄林梅, 张桂林, 王新余. 基于动态规划的红外运动小目标检测算法[J]. 红外激光与工程, 2004, 33(3): 303- 306.
[4] 蔡智富, 赵 坤, 杨莘元. 基于 DSP 与 FPGA 的红外起伏背景估计系统[J]. 应用科技, 2006, 33(8): 34- 36.
[5] 郭天天, 卢焕章. 基于 FPGA 的序列图像能量累加[J]. 信号处理, 2006, 22(5): 694- 696.
[6] 陈祖爵, 韩 云, 鞠时光. 运动估计算法设计及 FPGA 实现[J]. 江苏大学学报, 2007, 28(4): 340- 344.