一种基于格局的程序分析方法*

邢光荣 郑国梁 李宣东

(北京信息工程学院,北京 100101) (南京大学计算机系,南京 210008)

摘要 本文提出了一种自动程序分析方法,其基本思想是程序可以看成由一组基本成分根据特定的构造方式来构成,从而可以通过提供一组标准的分析方法,实现对程序的自动分析与理解.

关键词 格局,格局构造方法.

所谓程序分析,指通过对程序代码的阅读、分析,以便了解相应程序的结构及其所要完成的功能.程序分析可以看成是由程序代码到相应的程序规格说明的抽象,它是软件理解的基础,对软件维护和软件重用都有重要意义.

按照程序分析的不同目标,可以区分两种情况:程序的结构分析与程序的功能分析.程序结构分析的主要目的是确定系统的组成成份以及这些成份间的相互作用关系.程序功能分析指在结构分析的基础上,确定系统各成份的功能以至系统的功能.程序功能分析要比结构分析困难得多,目前尚无成熟的分析技术,主要依靠人工分析,这样做不仅效率低,而且容易出错.

本文将提出一种程序分析方法,其基本思想是:一个程序可划分为若干作用相对固定的基本成份,它们以某种固定的方式构成更高层的成份直至整个程序.我们把按这种方法分析得到的结果称为程序格局,相应地这种方法称为基于格局的程序分析方法.采用这种方法,既可以实现程序结构分析,又可以在一定程度上实现程序的功能分析.

1 程序分析的基本原理

在程序分析中,通常将一个程序分解成若干子程序,由对这些子程序的理解得出对整个程序的理解.因此,分析方法的关键在于如何划分程序.可以从3个方面来评价一种分析方法:首先,对于任意程序,能否方便地确定该程序的组成成份;其次,分解得来的成份是否比整个程序容易理解;第三,由对成份的理解能否方便地得到对整个程序的理解.

先看一个例子:

例 1:z:= 0;

* 本文1991-05-13收到,1992-03-11定稿

作者**邢光荣**,28岁,1990年硕士毕业于南京大学,主要研究领域为软件工程环境,软件理解与软件维护等.**郑国梁**,57岁,教授,主要研究领域为计算机软件,软件工程.**李宣东**,31岁,博士生,主要研究领域为软件工程.

本文通讯联系人:郑国梁,南京 210008,南京大学计算机系

```
for I in 1…N loop
  if A(I)>0 then
    z:= z + A(I);
  end if
end loop
```

对这段程序,最简单的方法是逐行分析,把每一程序行看成一个基本成份.显然,程序分解工作十分简单,各成份也较易理解.但是,由于成份间的作用关系十分复杂,很难由成份的理解得出整个程序的理解.

按照结构化程序设计的理论,程序可以看成是由一些基本成份按照顺序、条件及循环三种基本结构有层次地建立起来的.根据这一理论,可以将上例中的程序分解成赋值语句'z:=0'和循环两部分,循环的循环体为一个条件语句,其条件成立分枝为赋值语句'z:=z+A(I)'.

这样分解的结果,各成份都较易理解,并且对于顺序结构和条件结构,由成份的理解得到整体的理解也不困难.至于循环结构,循环体的分析可同样进行.问题在于如何由循环体的理解得到对循环的理解?由于一个循环通常包括三个部分:初始化、循环控制以及循环体,而这三部分并不都直接出现在循环结构中,对循环的理解十分困难.

事实上,一个循环可以看成是对循环体的若干次顺序执行的复合,如例1中的循环语句可分解如下:

```
I:=1;if A(I)>0 then z:=z+A(I) end if;
I:=2;if A(I)>0 then z:=z+A(I) end if;
……
I:=N;if A(I)>0 then z:=z+A(I) end if;
```

由此看出,上例中循环相当于两部分的复合:对循环控制变量进行的枚举和利用枚举结果进行运算的循环体.进一步可将上例分解成如下三个片段:

```
例2:for I in 1…N loop      if A(I)>0 then      z:=0;
    end loop                end if              z:=z+A(I)
      片段 A                  片段 B              片段 C
```

这里片段 A 相当于一个对整数序列1…N 进行枚举的过程;片段 B 选择 A 所产生的枚举序列中满足给定条件的元素;片段 C 计算由 B 所选择的序列之和.这样分解有两个优点:一、循环被分解成一些易于理解的固定类型片段;二、这些片段的复合逻辑上等价于顺序复合,因而易于从片段的理解得出循环的理解.在这种分解方法的启发下,我们得出基于格局的程序分析方法,其核心思想在于提出一组格局构造方法,根据它们构造相应程序的程序格局,完成对程序的分析理解.

2　格局构造方法

根据所要分析程序结构的不同,有两类格局构造方法:无循环程序的格局构造方法和循环程序的格局构造方法.

2.1 无循环程序的格局构造方法

无循环程序的格局构造方法有三种:谓词型、条件型及复合型,其中复合型可进一步区分为并列型和表达式型两种情况.

(1)谓词型

该方法用于构造非终结的测试格局,它有三个基本成份:条件测试、动作及汇聚,通常用来描述由简单测试构造复杂测试.这里,汇聚用来表示条件测试的不同路径执行之交汇点.

(2)条件型

该方法用于构造条件执行类格局,相当于通常的条件语句结构,它由三种基本成份组成:条件、操作以及汇聚.例如:

例3:if x<0 then

x:=-x;

end if

这里'x<0'为条件,赋值语句'x:=-x'为操作,汇聚含义同前.

(3)并列型

并列型格局构造方法用于将一些相互无作用的片段组合成格局.在这类格局中,基本成份间无数据流关系,可对它们分别进行分析.

(4)表达式型

在这类格局中,基本成份间存在一定的相互作用,即一个片段对另一片段存在数据依赖性.

2.2 循环程序格局构造方法

循环程序格局构造方法的思想是将循环分解成一些固定类型的成份,这些成份包括:基本循环、操作、判断及终止.

(1)基本循环

称循环中具有如下形式的部分为基本循环:

for I in 1…N loop

end loop

这实际上是一个循环体为空的循环语句,其作用在于产生一个临时的整数序列,至于该序列的使用情况则不加考虑.有一种特殊情形:

loop end loop

它相当于对整数集的枚举.

(2)操作

循环所要完成的特定计算由操作部分来指明.一般来说,可以把操作看成是利用基本循环所产生的时序值,完成特定的计算任务.操作通常由两部分组成:初始化和运算体.当循环仅包含单一操作时,运算体的识别相对简单.更一般的情况是循环中包含多个操作.

为方便对循环中多个操作的识别,我们给出如下限制:一操作的运算体可以使用循环外或该循环其它部分的数据,但不能使用另一操作所产生的数据.因此,求循环中无数据流到其它部分的最小代码集,可以确定循环中操作的运算体.

由于同一循环中不同操作之间无数据关联,因而删除其中某一操作,并不影响对循环中

其它成份的分析.事实上,可以把一个复杂循环分解成一个操作和一个相对简单的循环,从而简化循环的分析.

(3)判断

判断可以看成是一种特殊的操作,其运算体是一个条件结构,相应的真假分枝均为空(如例2中的片段 B).

判断利用基本循环产生的时序值作为输入,对它进行选择以提供给其它成份使用.与操作类似,判断可以从循环中分离出来,以简化循环的分析.

(4)终止

终止是循环的一个特殊成份,其作用是导致循环执行的结束.请看下例:

例4:I:=1;z:=0;

```
while I<=N do loop
    z:=z+A(I);
    I:=I+1;
end loop
```

可以将该程序分解成如下三个片段:

片段 A: while I<=N do

片段 B: loop end loop

片段 C: I:=1;z:=0

```
    z:=z+A(I);I:=I+1;
```

称循环中形如片段 A 的部分为循环的终止部分,它相当于一个条件判断.

3　基于格局的程序分析

3.1　格局

格局是程序的一种抽象表示,它仅描述程序内部的基本逻辑特征.在格局中,程序的控制流信息和数据流信息均被显式描述.

格局的基本构造单位是段,或称基本段.段对应程序的基本操作,它有若干个输入端口和输出端口,分别用来指明该单元的输入值和所产生的输出值.有两类基本段:运算和测试.运算又称简单段,它对应程序中的基本运算,有一个入口点和一个出口点.测试也称分岔段,相当于条件测试,有一个入口点和真假两个出口点.

组成格局的成份可以是基本段,也可以是中间子格局,由这些成份构成格局的方法即为格局构造方法.根据构成格局的构造方法的不同,格局中各成份所起的作用十分明确,成份之间的作用关系也很清楚.格局不仅描述了程序的组成成份,而且通过对各成份在格局中所起作用的描述,可以表达有关程序的功能.

3.2　分析过程

基于格局的程序分析过程实际上就是利用格局构造方法来构造相应程序的格局,这一过程包括四个步骤:语法语义分析、表层格局的生成、分组以及时序分解.

(1)语法语义分析

对程序进行语法分析，检查其语法正确性并将其转换成一种中间描述.在此基础上对程序进行控制流和数据流的分析，并将其中的复杂控制结构(如 case 结构)转换成基本控制结构.

(2)表层格局的生成

根据语法语义分析的结果，沿着程序的每一条控制路径进行处理，建立相应的片段及片段间的控制流、数据流关系.

这一阶段分析得到的结果称为表层格局，它描述了程序中所有的基本段以及基本段之间的数据流和控制流关系.

(3)分组

分组指通过在表层格局中插入适当的中间段，以实现格局的层次结构.中间段是根据格局构造方法引入的.

(4)时序分解

时序分解的主要工作是对循环程序进行分解，识别其中各成份，完成对整个循环的分析理解.

按照格局构造方法，循环将被分解为四种成份：基本循环、操作、测试以及循环终止.在这四种成份中，基本循环和循环终止成份的识别相对简单.由于测试可看成一类特殊的操作，这样问题归结为对循环中操作的识别.

操作由初始化和运算体组成.前面已经介绍，运算体可以通过求循环中无数据流到其它部分的最小代码集来确定.至于循环初始化部分，虽然它直接影响到循环的执行，但并不出现在循环之中.我们认为：循环的初始化部分是这样一些代码，它们存在数据流流向某个循环并且其数据流仅流向该循环，因而，可以通过分析数据流来识别循环的初始化部分.

3.3 利用格局构造方法的分析程序

基于前面介绍的方法，作者实现了一个针对 Ada 程序的分析程序，它是我们实现的程序员助手 NPA 的一个组成部分，可以将用户提供的 Ada 程序分析抽象，得出程序格局存放到格局库中，以便提供给 NPA 的编码程序生成具有类似结构的应用程序.例如，通过对财务部门周报表处理程序的分析，可以得出较通用的报表处理程序格局，以后可根据它生成相应的月报表处理程序.

总结：基于格局的程序分析思想是：程序由一些起固定作用的成份按特定的方式组合而成，因而可以找到一组标准方法来分析程序.按照这一思想，作者实现了一个分析程序，以实现对 Ada 程序的分析.

由于格局构造方法是在结构程序设计基础上提出的，其分析能力必然受到某些限制.事实上，对于非结构化程序，仅仅按照前面介绍的方法有可能无法进行分析.如何将格局构造方法扩充到非结构化程序的分析中，有待进一步研究.此外，限于篇幅，对如何分析递归程序也未作讨论.

参考文献

1 Waters R C. The programmer's apprentice: a session with KBEmacs. IEEE Trans. Soft. Eng., 1985; SE-11 (11).

2 Waters R C. The programmer's apprentice: knowledge based program editing. IEEE Trans. Soft. Eng., 1982; **SE-8(1)**.
3 Waters R C. A system for understanding mathematical fortran programs. Rep. MIT/AI Memo., 1976(**368**).
4 Waters R C. A methods for analyzing loop programs. IEEE Trans. Soft. Eng., 1979; **SE-5(5)**.
5 Waters R C. A methods for automatically analyzing programs. Proc. of 6th IJCAI, Aug. 1979.
6 郑国梁，邢光荣，李存珠. 程序员助手 NPA. 计算机学报，1992；**15(12)**：933.

A METHOD FOR PROGRAM ANALYSIS BASED ON PROGRAM PLAN

Xing Guangrong

(*Beijing Information Technology Institute, Beijing* 100101)

Zheng Guoliang and Li Xuandong

(*Department of Computer Science, Nanjing University, Nanjing* 210008)

Abstract This paper describes a method for automatically analyzing programs. Its implication is that a program consists of some elementary components which interact each other in stereotyped models, hence the authors may give some standard analyzing methods to analyze and understand programs automatically.

Key words Plan, plan building method.

第三届中国人工智能联合学术会议

征　文　通　知

由国防科工委系统工程研究所和北京航空航天大学联合承办的第三届中国人工智能联合学术会(CJCAI'94)将于1994年秋季在重庆北碚举行。为了交流两年来的研究成果，更进一步地探索和开拓有中国特色的人工智能的研究道路，使大会开得更有成效，学术气氛更浓，今诚恳邀请全国有关专家、学者、科研和工程技术人员积极撰写学术论文并参加学术活动。现将会议征文的有关事项通知如下：

一、征文范围

AI 的基础理论	各种基于知识的系统	机器学习与知识获取	AI 与并行处理
AI 的语言	智能辅助教学	模式识别	人工智能应用
AI 系统体系结构	自动推理	神经网络	其它有关 AI 的内容
知识表示			

二、征文时间：征文截止时间：1994－06－30，录用通知发出时间：1994－07－15

三、注意事项：1. 来稿必须是可以公开发表的学术论文，不涉及国家机密内容；2. 来稿应是未在公开学术刊物上正式发表的论文；3. 论文应字迹清晰，插图完整，字数限制在6000字以内，并应有300字以内的摘要；参考文献应注明出处；论文要一式两份；4. 全部论文将由大会程序委员会负责审稿，论文一经录用，将发录用通知；5. 全部论文将由大会统一编入"论文集"；6. 因工作人员有限，来稿不论入选与否，概不退稿，请作者自留底稿；7. 来稿请寄：**北京**9702**信箱**19**号**(**邮编**100101)**袁世崇收**，联系电话：6756510；8. 会议具体时间、地点另行通知。