



普通高等教育“十一五”规划教材

《编译原理》

习题解析与上机指导

(第三版)

主编 王 磊 胡元义

00010001010 10101100
010110 00010001010101100
00010001010 10101100010110
0001000101010101100 010101010
10001010010000
10101010010111 010110 01011111010010
0001000101010101100 010101010 10001010010000
00010001010 10101100 010110 000100010101010110
00010001010 10101100010110
000100010101010101100 010101010 10001010010000
10101010010111 010110 01011111010010
0001000101010101100 010101010 1000101001000000010
10101100 010110 0001000101010101100
00010001010 10101100010110
000100010101010101100 010101010 10001010010000
10101010010111 010110 01011111010010
0001000101010101100 010101010
100010100100000010001010 10101100 010110
0001000101010101100
00010001010 10101100010110
000100010101010101100 010101010
10001010010000
000100010101010101100 010101010
010 10101100



科学出版社
www.sciencep.com

普通高等教育“十一五”规划教材

《编译原理》
习题解析与上机指导
(第三版)

科学出版社

北 京

内 容 简 介

本书是《编译原理》(王磊、胡元义主编,科学出版社出版,第三版)一书配套的习题解析与上机指导教材,也可单独使用。本书的习题解析部分对《编译原理》中的习题进行了深入、细致的分析和解答,为读者熟练掌握编译原理知识、抓住重点、突破难点提供了有益的帮助。本书的上机指导部分包括由高级语言到中间语言、由汇编语言到机器语言的翻译,使编译的主要翻译阶段和环节都能微观且实时地显示出来,较好地解决了编译原理的理论与实践的衔接问题。此外,我们结合自己开发的 8086/8088 小汇编指令到机器代码的翻译成果,将 8086/8088 汇编指令如何翻译成机器代码的方法引入到书中,有利于读者了解低级语言的翻译过程和实现方法。

本书可作为计算机专业的本科教材,也可作为计算机软件工程人员的参考资料。

图书在版编目(CIP)数据

《编译原理》习题解析与上机指导 / 王磊,胡元义主编. —3 版. —北京:科学出版社,2009

普通高等教育“十一五”规划教材

ISBN 978-7-03-025125-1

I. 编… II. ①王…②胡… III. 编译程序—程序设计—高等学校—教学参考资料 IV. TP314

中国版本图书馆 CIP 数据核字(2009)第 132687 号

责任编辑:贾瑞娜 / 责任校对:陈玉凤

责任印制:张克忠 / 封面设计:耕者设计工作室

科学出版社出版

北京东黄城根北街 16 号

邮政编码:100717

<http://www.sciencep.com>

骏立印刷厂印刷

科学出版社发行 各地新华书店经销

*

2009 年 8 月第 三 版 开本: B5 (720×1000)

2009 年 8 月第一次印刷 印张: 14 3/4

印数: 1—3 500 字数: 286 000

定价: 23.00 元

(如有印装质量问题,我社负责调换)

前 言

编译原理是计算机专业的一门核心课程,在计算机本科教学中占有十分重要的地位。由于编译原理课程具有很强的理论性和实践性,因而学生在学习时普遍感到内容抽象,不易理解,掌握起来难度较大。本书通过习题解析的方式来帮助读者理解编译技术的原理和概念,掌握编译原理的相关方法,提高分析问题与解决问题的能力。本书的上机指导部分则给读者提供了一个完整的小型编译程序,以便读者上机实践,较好地解决了编译原理与实践的衔接问题,使读者对编译原理有一个形象、直观和透彻的认识及感受,以便深入了解和掌握编译原理的内容及实现方法。

本书是与编者在科学出版社出版的《编译原理》(第三版)一书相配套的习题解析与上机指导教材,它也可以与目前各种编译原理教材配套使用。本书分为两篇,第一篇为编译原理习题解析部分。为了便于读者正确理解编译原理的概念,掌握解题方法,本篇对《编译原理》(第三版)一书中各章的习题都给出了详尽的解题过程以及引用到的概念、原理和公式的出处;对有代表性的习题和疑难习题也给出了详细的分析和解答。此外,对某些习题,本书还给出了一些新的解题思路和方法。本书的第二篇是编译上机指导部分,给出了一个完整的小型编译程序,该程序涵盖了编译原理的词法分析、语法分析、中间代码生成等各阶段的内容。此外,还给出了 8086/8088 汇编语言到机器语言的翻译程序。本书中的小型编译程序可接受本书中文法规定的高级语言程序,并将其翻译成四元式代码形式的中间语言程序,并且使编译的主要翻译阶段和环节都能微观且实时地显示出来,有利于读者深入了解编译的内部过程和实现细节,并为读者开拓了进一步学习和运用编译原理的视野。对目标代码生成,国内的编译教材只是笼统地介绍了从中间代码到假想机汇编这一级的翻译,究竟计算机是如何实现将汇编语言翻译成可执行的机器代码却均无介绍。我们结合自己开发的 8086/8088 小汇编指令到机器代码的翻译成果,将 8086/8088 汇编指令如何翻译成机器代码的方法引入到书中,有利于读者了解低级语言的翻译过程和实现方法。

由于编者水平所限,本书难免存在差错和不足,敬请广大读者批评指正。

编 者

2009 年 5 月

目 录

前言

第一篇 习题解析

第 1 章 绪论	3
第 2 章 词法分析	5
第 3 章 语法分析	15
第 4 章 语义分析和中间代码生成	58
第 5 章 代码优化	67
第 6 章 目标程序运行时存储空间的组织	78
第 7 章 目标代码生成	85
第 8 章 符号表与错误处理	90

第二篇 上机指导

第 9 章 小型编译程序介绍	97
9.1 小型编译程序结构	97
9.2 小型编译程序关于高级语言的规定	98
9.3 小型编译程序关于单词的内部定义	99
9.4 小型编译程序的 LR 分析表	101
9.5 小型编译程序执行过程及实例分析	103
第 10 章 上机实验内容	109
10.1 实验一 编译程序的分析与验证	109
10.2 实验二 算术表达式的扩充	110
10.3 实验三 添加新的程序语句(一)	110
10.4 实验四 添加新的程序语句(二)	111
10.5 编译原理课程设计	112
第 11 章 高级语言到四元式的编译程序	114
第 12 章 8086/8088 小汇编的设计与实现	141
12.1 汇编指令系统的分析	141
12.2 8086/8088 小汇编的设计与实现	149

12.3 8086/8088 小汇编实验	158
12.4 8086/8088 小汇编程序	161
参考文献	219
附录 1 8086/8088 指令码汇总表	220
附录 2 8086/8088 指令编码空间表	225

第一篇

习题解析

第1章 绪 论

1.1 完成下列选择题:

(1) 构造编译程序应掌握_____。

- a. 源程序
- b. 目标语言
- c. 编译方法
- d. 以上三项都是

(2) 编译程序绝大多数时间花在_____上。

- a. 出错处理
- b. 词法分析
- c. 目标代码生成
- d. 表格管理

(3) 编译程序是对_____。

- a. 汇编程序的翻译
- b. 高级语言程序的解释执行
- c. 机器语言的执行
- d. 高级语言的翻译

【解答】 (1) d (2) d (3) d

1.2 计算机执行用高级语言编写的程序有哪些途径? 它们之间的主要区别是什么?

【解答】 计算机执行用高级语言编写的程序主要有两种途径:解释和编译。

在解释方式下,翻译程序事先并不采用将高级语言程序全部翻译成机器代码程序,然后执行这个机器代码程序的方法,而是每读入一条源程序的语句,就将其解释(翻译)成对应其功能的机器代码语句串并执行,而所翻译的机器代码语句串在该语句执行后并不保留,然后再读入下一条源程序语句,并解释执行。这种方法是按源程序中语句的动态执行顺序逐句解释(翻译)执行的,如果一语句处于一循环体中,则每次循环执行到该语句时,都要将其翻译成机器代码后再执行。

在编译方式下,高级语言程序的执行是分两步进行的:第一步首先将高级语言程序全部翻译成机器代码程序,第二步才是执行这个机器代码程序。因此,编译对源程序的处理是先翻译,后执行。

从执行速度上看,编译型的高级语言比解释型的高级语言要快,但解释方式下的人机界面比编译型的好,便于程序调试。

这两种途径的主要区别在于:解释方式下不生成目标代码程序,而编译方式下生成目标代码程序。

1.3 请画出编译程序的总框图。如果你是一个编译程序的总设计师,设计编译程序时应当考虑哪些问题?

【解答】 编译程序总框图如图 1-1 所示。

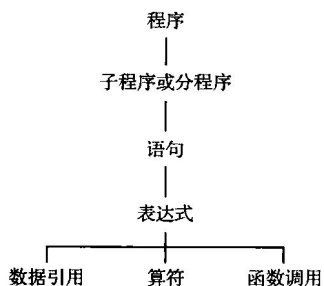


图 1-1 编译程序总框图

作为一个编译程序的总设计师,首先要深刻理解被编译的源语言其语法及语义;其次,要充分掌握目标指令的功能及特点,如果目标语言是机器指令,还要搞清楚机器的硬件结构以及操作系统的功能;第三,对编译的方法及使用的软件工具也必须准确化。总之,总设计师在设计编译程序时必须估量系统功能要求、硬件设备及软件工具等诸因素对编译程序构造的影响等。

第2章 词法分析

2.1 完成下列选择题:

- (1) 词法分析器的输出结果是_____。
- a. 单词的种别编码
 - b. 单词在符号表中的位置
 - c. 单词的种别编码和自身值
 - d. 单词自身值
- (2) 正规式 M_1 和 M_2 等价是指_____。
- a. M_1 和 M_2 的状态数相等
 - b. M_1 和 M_2 的有向边条数相等
 - c. M_1 和 M_2 所识别的语言集相等
 - d. M_1 和 M_2 状态数和有向边条数相等
- (3) DFA M(图 2-1)接受的字集为_____。

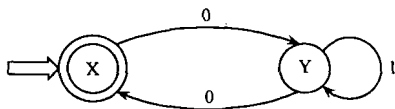


图 2-1 习题 2.1 的 DFA M

- a. 以 0 开头的二进制数组成的集合
- b. 以 0 结尾的二进制数组成的集合
- c. 含奇数个 0 的二进制数组成的集合
- d. 含偶数个 0 的二进制数组成的集合

【解答】 (1) c (2) c (3) d

2.2 什么是扫描器? 扫描器的功能是什么?

【解答】 扫描器就是词法分析器,它接受输入的源程序,对源程序进行词法分析并识别出一个个单词符号,其输出结果是单词符号,供语法分析器使用。通常把词法分析器作为一个子程序,每当词法分析器需要一个单词符号时就调用这个子程序。每次调用时,词法分析器就从输入串中识别出一个单词符号交给语法分析器。

2.3 设 $M = (\{x, y\}, \{a, b\}, f, x, \{y\})$ 为一非确定的有限自动机,其中 f 定义如下:

$$\begin{aligned} f(x, a) &= \{x, y\} & f(x, b) &= \{y\} \\ f(y, a) &= \Phi & f(y, b) &= \{x, y\} \end{aligned}$$

试构造相应的确定有限自动机 M' 。

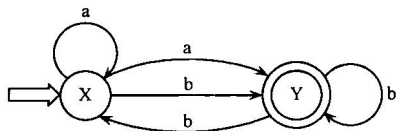


图 2-2 习题 2.3 的 NFA M

【解答】 对照自动机的定义 $M = (S, \Sigma, f, S_0, Z)$, 由 f 的定义可知 $f(x, a)$ 、 $f(y, b)$ 均为多值函数, 因此 M 是一非确定有限自动机。

先画出 NFA M 相应的状态图, 如图 2-2 所示。

用子集法构造状态转换矩阵, 如表 2-1 所示。

表 2-1 状态转换矩阵

I	I_a	I_b
$\{x\}$	$\{x, y\}$	$\{y\}$
$\{y\}$	—	$\{x, y\}$
$\{x, y\}$	$\{x, y\}$	$\{x, y\}$

将转换矩阵中的所有子集重新命名, 形成表 2-2 所示的状态转换矩阵。

表 2-2 状态转换矩阵

状态 \ f 字符	字符	
	a	b
0	2	1
1	—	2
2	2	2

即得到 $M' = (\{0, 1, 2\}, \{a, b\}, f, 0, \{1, 2\})$, 其状态转换图如图 2-3 所示。

将图 2-3 所示的 DFA M' 最小化。首先, 将 M' 的状态分成终态组 $\{1, 2\}$ 与非终态组 $\{0\}$ 。其次, 考察 $\{1, 2\}$, 由于 $\{1, 2\}_a = \{1, 2\}_b = \{2\} \subset \{1, 2\}$, 所以不再将其划分, 也即整个划分只有两组:

$\{0\}$ 和 $\{1, 2\}$ 。令状态 1 代表 $\{1, 2\}$, 即把原来到达 2 的弧都导向 1, 并删除状态 2。最后, 得到如图 2-4 所示的化简了的 DFA M' 。

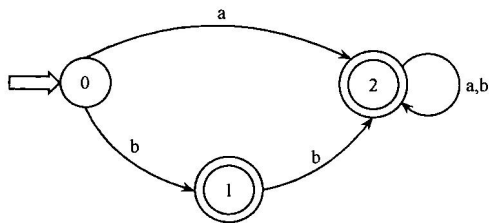
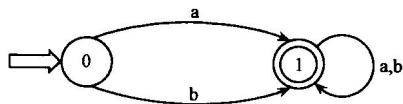
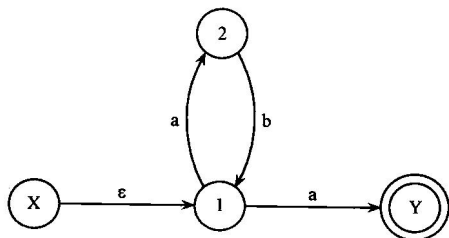
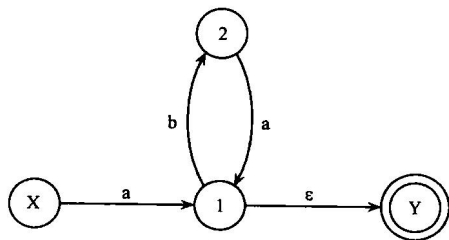


图 2-3 习题 2.3 的 DFA M'

图 2-4 图 2-3 化简后的 DFA M'

2.4 正规式 $(ab)^*a$ 与正规式 $a(ba)^*$ 是否等价? 请说明理由。

【解答】正规式 $(ab)^*a$ 对应的 NFA 如图 2-5 所示,正规式 $a(ba)^*$ 对应的 NFA 如图 2-6 所示。

图 2-5 正规式 $(ab)^*a$ 对应的 NFA图 2-6 正规式 $a(ba)^*$ 对应的 DFA

用子集法将图 2-5 和图 2-6 分别确定化为如图 2-7(a)和(b)所示的状态转换矩阵,它们最终都可以得到最简 DFA,如图 2-8 所示。因此,这两个正规式等价。

I	I_a	I_b
$\{X, 1\}$	$\{2, Y\}$	—
$\{2, Y\}$	—	$\{1\}$
$\{1\}$	$\{2, Y\}$	—

(a)

I	I_a	I_b
$\{X\}$	$\{1, Y\}$	$\{2\}$
$\{1, Y\}$	—	—
$\{2\}$	$\{1, Y\}$	—

(b)

图 2-7 图 2-5 和图 2-6 确定化后的状态转换矩阵

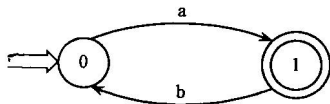
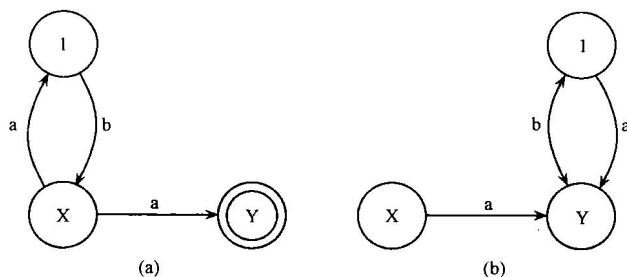


图 2-8 最简 NFA

实际上,当闭包 $*$ 取 0 时,正规式 $(ab)^*a$ 与正规式 $a(ba)^*$ 的由初态 X 到终态 Y 之间仅存在一条 a 弧。由于 $(ab)^*$ 在 a 之前,故描述 $(ab)^*$ 的弧应在初态节点 X 上;而 $(ba)^*$ 在 a 之后,故 $(ba)^*$ 对应的弧应在终态节点 Y 上。因此 $(ab)^*a$ 和 $a(ba)^*$ 所对应的 NFA 也可分别描述为如图 2-9(a)和(b)所示的形式,它们确定化并化简后仍可得图 2-8 所示的最简 NFA。

图 2-9 $(ab)^*a$ 和 $a(ba)^*$ 分别对应的 NFA

2.5 设有 $L(G) = \{a^{2n+1}b^{2m}a^{2p+1} \mid n \geq 0, p \geq 0, m \geq 1\}$;

(1) 给出描述该语言的正规表达式;

(2) 构造识别该语言的确有限自动机(可直接用状态图形式给出)。

【解答】该语言对应的正规表达式为 $a(aa)^*bb(bb)^*a(aa)^*$, 正规表达式对应的 NFA 如图 2-10 所示。

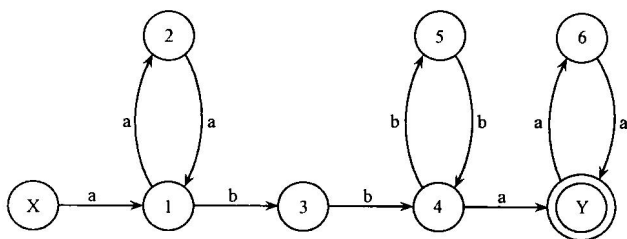


图 2-10 习题 2-5 的 NFA

用子集法将图 2-10 确定化, 如图 2-11 所示。

I	I_a	I_b
$\{X\}$	$\{1\}$	—
$\{1\}$	$\{2\}$	$\{3\}$
$\{2\}$	$\{1\}$	—
$\{3\}$	—	$\{4\}$
$\{4\}$	$\{Y\}$	$\{5\}$
$\{5\}$	—	$\{4\}$
$\{Y\}$	$\{6\}$	—
$\{6\}$	$\{Y\}$	—

重新命名

S	a	b
0	1	—
1	2	3
2	1	—
3	—	4
4	7	5
5	—	4
7	6	—
6	7	—

图 2-11 习题 2.5 的状态转换矩阵

由图 2-11 重新命名后的状态转换矩阵可化简为(也可由最小化方法得到)

$\{0,2\} \{1\} \{3,5\} \{4,6\} \{7\}$

按顺序重新命名为 0、1、2、3、4 后得到最简的 DFA,如图 2-12 所示。

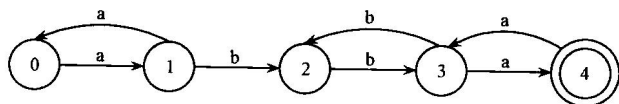


图 2-12 习题 2.5 的最简 DFA

2.6 有语言 $L = \{w \mid w \in (0,1)^+, \text{ 并且 } w \text{ 中至少有两个 } 1, \text{ 又在任何两个 } 1 \text{ 之间有偶数个 } 0\}$, 试构造接受该语言的确定有限状态自动机(DFA)。

【解答】 对于语言 L , w 中至少有两个 1, 且任意两个 1 之间必须有偶数个 0; 也即在第一个 1 之前和最后一个 1 之后, 对 0 的个数没有要求。据此我们求出 L 的正规式为 $0^*1(00(00)^*1)^*00(00)^*10^*$, 画出与正规式对应的 NFA, 如图 2-13 所示。

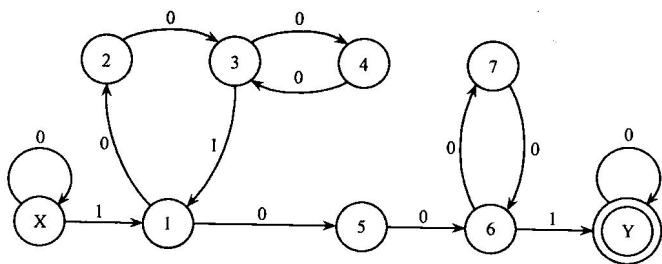


图 2-13 习题 2.6 的 NFA

用子集法将图 2-13 的 NFA 确定化, 如图 2-14 所示。

I	I_0	I_1
$\{X\}$	$\{X\}$	$\{1\}$
$\{1\}$	$\{2,5\}$	—
$\{2,5\}$	$\{3,6\}$	—
$\{3,6\}$	$\{4,7\}$	$\{1,Y\}$
$\{4,7\}$	$\{3,6\}$	—
$\{1,Y\}$	$\{2,5,Y\}$	—
$\{2,5,Y\}$	$\{3,6,Y\}$	—
$\{3,6,Y\}$	$\{4,7,Y\}$	$\{1,Y\}$
$\{4,7,Y\}$	$\{3,6,Y\}$	—

重新命名

S	0	1
0	0	1
1	2	—
2	3	—
3	4	5
4	3	—
5	6	—
6	7	—
7	8	5
8	7	—

图 2-14 习题 2.6 的状态转换矩阵

由图 2-14 可看出非终态 2 和 4 的下一状态相同,终态 6 和 8 的下一状态相同,即得到最简状态为

$$\{0\}, \{1\}, \{2, 4\}, \{3\}, \{5\}, \{6, 8\}, \{7\}$$

按顺序重新命名为 0、1、2、3、4、5、6,则得到最简 DFA,如图 2-15 所示。

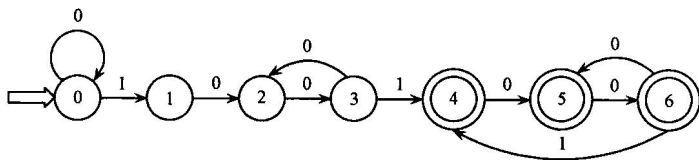


图 2-15 习题 2.6 的最简 DFA

2.7 已知正规式 $((a|b)^* | aa)^* b$ 和正规式 $(a|b)^* b$:

- (1) 试用有限自动机的等价性证明这两个正规式是等价的;
- (2) 给出相应的正规文法。

【解答】 (1) 正规式 $((a|b)^* | aa)^* b$ 对应的 NFA 如图 2-16 所示。

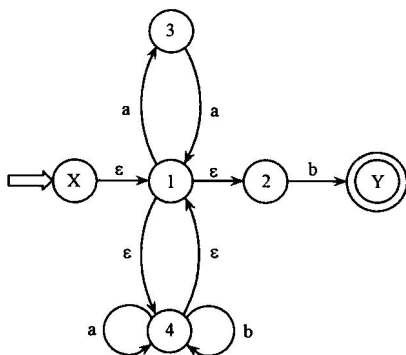


图 2-16 正规式 $((a|b)^* | aa)^* b$ 对应的 NFA

用子集法将图 2-16 所示的 NFA 确定化为 DFA,如图 2-17 所示。

I	I_a	I_b	重新命名 →	S	a	b
$\{X, 1, 2, 4\}$	$\{1, 2, 3, 4\}$	$\{1, 2, 4, Y\}$		1	2	3
$\{1, 2, 3, 4\}$	$\{1, 2, 3, 4\}$	$\{1, 2, 4, Y\}$		2	2	3
$\{1, 2, 4, Y\}$	$\{1, 2, 3, 4\}$	$\{1, 2, 4, Y\}$		3	2	3

图 2-17 图 2-16 确定化后的状态转换矩阵

由于对非终态的状态 1、2 来说,它们输入 a、b 的下一状态是一样的,故状态 1 和状态 2 可以合并,将合并后的终态 3 命名为 2,则得到表 2-3(注意,终态和非终态即使输入 a、b 的下一状态相同也不能合并)。

由此得到最简 DFA,如图 2-18 所示。

表 2-3 合并后的状态转换矩阵

S	a	b
1	1	2
2	1	2

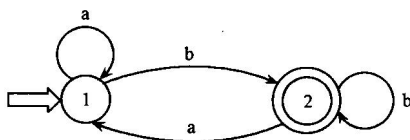
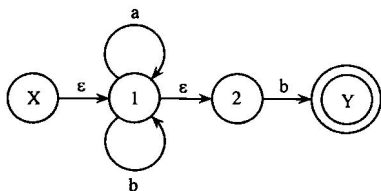


图 2-18 习题 2.7 的最简 DFA

正规式 $(a|b)^*b$ 对应的 NFA 如图 2-19 所示。

图 2-19 正规式 $(a|b)^*b$ 对应的 NFA

用子集法将图 2-19 所示的 NFA 确定化为如图 2-20 所示的状态转换矩阵。

I	I_a	I_b
$\{X, 1, 2\}$	$\{1, 2\}$	$\{1, 2, Y\}$
$\{1, 2\}$	$\{1, 2\}$	$\{1, 2, Y\}$
$\{1, 2, Y\}$	$\{1, 2\}$	$\{1, 2, Y\}$

重新命名

S	a	b
1	2	3
2	2	3
3	2	3

图 2-20 图 2-19 确定化后的状态转换矩阵

比较图 2-20 与图 2-17,重新命名后的转换矩阵是完全一样的,也即正规式 $(a|b)^*b$ 可以同样得到化简后的 DFA 如图 2-18 所示。因此,两个自动机完全一样,即两个正规文法等价。

(2) 对图 2-18,令 A 对应状态 1, B 对应状态 2,则相应的正规文法 $G[A]$ 为

$$G[A]: A \rightarrow aA | bB | b$$

$$B \rightarrow aA | bB | b$$

$G[A]$ 可进一步化简为 $G[S]: S \rightarrow aS | bS | b$ (非终结符 B 对应的产生式与 A 对应的产生式相同,故两非终结符等价,即可合并为一个产生式)。