

利用等价类构造有限状态自动机

蒋龙龙 陈文宇

(重庆邮电学院光电工程学院 重庆 400065) (电子科技大学计算机科学与工程学院 成都 610054)

摘要 一类语言由任意字母表上的某种进制的数字串构成,要求该语言中的所有数字串能够整除 N ;构造有限状态自动机识别该类语言是困难的,本文提出了根据等价类构造一类有限状态自动机的方法。该方法可以针对所有字母表和所有进制的数字串构成的语言,而且满足语言中的所有数字串能够整除任意正整数 N 。该方法实用、简便。

关键词 等价关系,等价类,有限状态自动机

Using the Equivalence Class to Make the Automaton

JIANG Long-Long CHEN Wen-Yu

(College of Electronic Engineering, CQUPT of China, Chongqing 400065)

(College of Computer Science and Engineering, UESTC of China, Chengdu 610054)

Abstract Some languages are made of the digit strings which can be divided by integer numb N . It is diffict to make the automaton to accept the language. Based on the equivalence class, we present the method of using equivalence to make the automaton. It can be used with the language of all alphabet and all system, and every digit string in language can divided by any integer numb N . It is convenient and usefull.

Keywords Equivalence relation, Equivalence class, Finite state automaton

1 等价关系和等价类

一些没有重复的对象的全体称为集合,而这些被包含的对象称为该集合的元素。集合中元素可以按任意的顺序进行排列。

定义 1 笛卡儿乘积的定义

集合 A 和 B 的笛卡儿乘积使用 $A \times B$ 表示,它是集合 $\{(a, b) \mid a \in A \text{ 且 } b \in B\}$ 。

定义 2 关系的定义

设 A 和 B 为两个集合,则由 A 到 B 的关系是 $A \times B$ 的某个子集。若 $A=B$,则称为 A 上的关系。若 R 为由 A 到 B 的关系,当 (a, b) 在 R 内时,可记为 aRb 。

定义 3 等价关系的定义

设 R 是集合 A 上的关系,那么

如果对 A 中的任一元素 a ,都有 aRa ,则称 R 为自反的。

如果对 A 中的任何元素 a 和 b ,从 aRb 能够推出 bRa ,则称 R 为对称的。

如果对 A 中的任何元素 a, b 和 c ,从 aRb 和 bRc 能够推出 aRc ,则称 R 为传递的。

如果关系 R 同时是自反的、对称的和传递的,则称之为等价关系。

等价关系的一个重要性质为:集合 A 上的一个等价关系 R 可以将集合 A 划分为若干个互不相交的子集,称为等价类。对 A 中的每个元素 a ,使用 $[a]$ 表示 a 的等价类,即 $[a] = \{b \mid aRb\}$ 。等价关系 R 将集合 A 划分成的等价类的数目,称为该等价关系的指数。

考虑非负整数集合 N 上的模 3 同余关系 $R, R = \{(a, b) \mid a, b \in N, \text{且 } a \bmod 3 = b \bmod 3\}$ 。那么,集合 $\{0, 3, 6, \dots, 3n,$

$\dots\}$ 形成一个等价类,记为 $[0]$;集合 $\{1, 4, 7, \dots, 3n+1, \dots\}$ 形成一个等价类,使用 $[1]$ 表示。集合 $\{2, 5, 8, \dots, 3n+2, \dots\}$ 形成另一个等价类,记为 $[2]$ 。

$N = [0] \cup [1] \cup [2]$; R 的指数为 3。

2 有限状态自动机

定义 4 有限状态自动机的定义

字母表 Σ 上的有限状态接收机 (FSAM) 是一个五元式, $FSAM = (Q, \Sigma, \delta, q_0, F)$,

其中:

Q 是一个有限状态的集合;

Σ 是字母表,也就是输入带上的字符的集合;

$q_0 \in Q$ 是开始状态;

FCQ 是接收状态(终止状态)集合;

δ 是 $Q \times \Sigma \rightarrow Q$ 的状态转换函数,即 $\delta(q, x) = q'$;代表自动机在状态 q 时,扫描字符 x 后到达状态 q' 。

有限状态自动机 $FSAM = (Q, \Sigma, \delta, q_0, F)$, 对于状态 $q, q \in Q$;能将有限状态自动机从开始状态转换到 q 状态的所有字符串的集合为:

$set(q) = \{w \mid w \in \Sigma^*, \delta(q_0, w) = q\}$

一般地,对于任意一个有限状态自动机 $FSAM = (Q, \Sigma, \delta, q_0, F)$,可以定义关系 R_M ,若 $x, y \in \Sigma^*$,则 $xR_M y \Leftrightarrow x \in set(q)$ 且 $y \in set(q)$,其中: $q \in Q$ 。

该关系是集合 Σ^* 上的一个等价关系,利用该关系,可以将 Σ^* 划分为不多于 $|Q|$ 个的等价类。

有限状态自动机可以按照语言的特点给出字母表 Σ^* 的一个划分,这种划分相当于 Σ^* 上的一个等价分类,有限状态自动机的每个状态实际上对应着一个等价类。所以,利用一

个状态去表示一个等价类是考虑问题(构造有限状态自动机)的一条有效思路。

3 经典问题

问题 1: 构造有限状态自动机, 识别 $\{0, 1, 2\}$ 上的语言, 该语言的每个字符串代表的十进制整数能整除 3。

分析: 若一个十进制整数的所有位的数字的和能够整除 3, 那么, 这个十进制整数本身就能够整除 3;

一个十进制整数除以 3, 余数只能是 1、2 和 0 (余数为 0, 则表示该数能够整除 3);

将整数当作一个字符串, 从左到右逐一地读入每个字符;

使用 3 个状态分别代表已经读入的数字的和除以 3 的不同的余数的情况:

q_0 : 已经读入的数字的和除以 3, 余数为 0 的输入串的等价类;

q_1 : 已经读入的数字的和除以 3, 余数为 1 的输入串的等价类;

q_2 : 已经读入的数字的和除以 3, 余数为 2 的输入串的等价类;

则扫描子串 w 后, 处于某个状态, 读入当前数字, 有限自动机的状态转换情况为:

q_0 : 在此状态读入 0, 引导自动机到达下一状态的输入串为 $w0$, $w0$ 的各位数字和也属于 q_0 对应的等价类。所以, 自动机在 q_0 状态读入 0, 应该继续保持 q_0 状态;

在此状态读入 1, 引导自动机到达下一状态的输入串为 $w1$, $w1$ 的各位数字和属于 q_1 对应的等价类。所以, 自动机在 q_0 状态读入 1, 应该到达 q_1 状态;

在此状态读入 2, 引导自动机到达下一状态的输入串为 $w2$, $w2$ 的各位数字和属于 q_2 对应的等价类。所以, 自动机在 q_0 状态读入 2, 应该到达 q_2 状态;

q_1 : 在此状态读入 0, 引导自动机到达下一状态的输入串为 $w0$, 则 $w0$ 的各位数字和也属于 q_1 对应的等价类。所以, 自动机在 q_1 状态读入 0, 应该保持 q_1 状态;

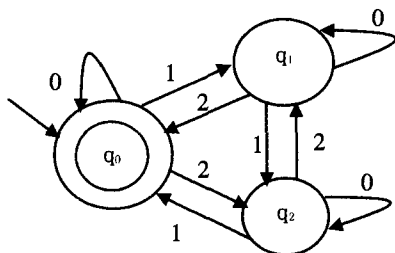
在此状态读入 1, 引导自动机到达下一状态的输入串为 $w1$, $w1$ 的各位数字和属于 q_2 对应的等价类。所以, 自动机在 q_1 状态读入 1, 应该到达 q_2 状态;

在此状态读入 2, 引导自动机到达下一状态的输入串为 $w2$, $w2$ 的各位数字和属于 q_0 对应的等价类。所以, 自动机在 q_1 状态读入 2, 应该到达 q_0 状态;

q_2 : 在此状态读入 0, 引导自动机到达下一状态的输入串为 $w0$, 则 $w0$ 的各位数字和也属于 q_2 对应的等价类。所以, 自动机在 q_2 状态读入 0, 应该保持 q_2 状态;

在此状态读入 1, 引导自动机到达下一状态的输入串为 $w1$, $w1$ 的各位数字和属于 q_0 对应的等价类。所以, 自动机在 q_2 状态读入 1, 应该到达 q_0 状态;

在此状态读入 2, 引导自动机到达下一状态的输入串为 $w2$, $w2$ 的各位数字和属于 q_1 对应的等价类。所以, 自动机在 q_2 状态读入 2, 应该到达 q_1 状态。



(注意: 接收的串包括以 0 开始的数字串; 还能够接收空串。此问题在此不讨论)

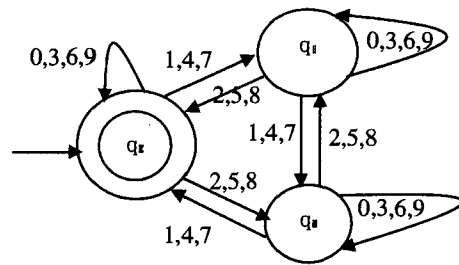
问题 2: 构造有限状态自动机 M , 识别 $\{0, 1, 2, 4, 5, 6, 7, 8, 9\}$ 上的语言, 该语言的每个字符串代表的十进制整数能整除 3。

分析: 一个十进制整数除以 3, 余数只能是 1、2 和 0; 仍然只使用 3 个状态分别代表已经读入的数字的和除以 3 的不同的余数的情况:

q_0 : 已经读入的数字的和除以 3, 余数为 0 的输入串的等价类;

q_1 : 已经读入的数字的和除以 3, 余数为 1 的输入串的等价类;

q_2 : 已经读入的数字的和除以 3, 余数为 2 的输入串的等价类。



4 一般问题

对于二进制整数, 或十进制整数, 要求代表的数字能整除 N , 没有具体的数学算法。对于一类有限状态自动机, 可以描述为:

构造有限状态自动机 $FSAM$, 识别字母表 $\Sigma = \{0, 1, 2, \dots, M-1\}$ 上的语言, 该语言的每个字符串当成 M 进制数时, 代表的数字能整除 N 。

分析: 将不同进制的数转换为十进制数后, 除以 N 的余数只能为 $0, 1, 2, 3, \dots$ 和 $N-1$, 使用 N 个状态分别代表已经读入的串代表的数字除以 N 的不同的余数的等价类:

q_0 : 已经读入的数除以 N , 余数为 0 的输入串的等价类; 该类数为 $N * n + 0$;

q_1 : 已经读入的数除以 N , 余数为 1 的输入串的等价类; 该类数为 $N * n + 1$;

q_2 : 已经读入的数除以 N , 余数为 2 的输入串的等价类; 该类数为 $N * n + 2$;

...

q_{N-1} : 已经读入的数除以 N , 余数为 $N-1$ 的输入串的等价类; 该类数为 $N * n + N-1$;

注意: 因为不能接收空串, 所以还需要一个开始状态 q_s 。

q_s : 在开始状态读入 x 时, 进入对应状态 q_x ;

q_i : 对应已经读入的数 w 除以 N , 余数为 i 的输入串的等价类; 该类数为 $N * n + i$;

当前读入的字符为 x ; 则 wx 表示的十进制数为:

$$M * (N * n + i) + x \\ = N * M * n + M * i + x$$

该数对于 N 取余数就是 $M * i + x$ 对于 N 的余数, 若该余数为 j , 则相应的状态就应该从 q_i 变换为 q_j 。

实例: 构造有限状态自动机 M , 识别 $\{1, 2, 3\}$ 上的语言, 该语言的每个字符串当成十进制数时, 代表的数字能整除 4。

(下转第 277 页)

图7表明了系统的非稳定性对于自愈率 λ_4 和故障恢复率 λ_{11} 的依赖性。

结论 本文综合考虑了容错系统的硬件故障和老化现象,提出了将传统的冗余策略和软件抗衰策略相结合的技术应用在容错系统中,并分析了该系统的状态变化,给出了系统的时间颜色 Petri 网(timed-CPN)模型。在结合数据分析容错系统的行为后,得出了系统非稳定概率和模块自愈时间间隔及故障恢复率之间的关系。

参考文献

- 1 Castelli V, Harper R E, Heidelberger P, et al. Proactive management of software aging. IBM J. Research & Development, 2001, 45:311~332

- 2 Marshall E. Fatal error: how Patriot overlooked a scud. Science, 1992, 255:1347
- 3 Huang Y, Kintala C, Kolettin N, et al. Software rejuvenation: analysis, module and applications. In: Proc. 25th Int'l Symp. on Fault Tolerant Computing, IEEE CS Press, 1995. 381~390
- 4 Shereshevsky M, Cukic B, Crowel J, et al. Software aging and multifractality of memory resources. In: Proc. Int'l Conf. on Dependable Systems and Networks, IEEE CS Press, 2003. 721~730
- 5 Garg S, Telek M, Puliafito A. Analysis of software rejuvenation using Markov regenerative stochastic Petri net. In: Proc. 6th Int'l Symp. on Software Reliab. Eng. IEEE CS Press, 1995. 24~27
- 6 Genrich H J. Equivalence transformation of prt-net. In: Proc. of 9th European workshop on Application and Theory of Petri Net, Venice, Italy, 1988

(上接第273页)

分析:除以4的余数只能为0、1、2和3,使用4个状态分别代表已经读入的数字的和除以4的不同的余数的等价类:

q_0 :已经读入的数字的和除以4,余数为0的输入串的等价类;

q_1 :已经读入的数字的和除以4,余数为1的输入串的等价类;

q_2 :已经读入的数字的和除以4,余数为2的输入串的等价类;

q_3 :已经读入的数字的和除以4,余数为3的输入串的等价类;

因为不能接收空串,所以还需要一个开始状态 q_s 。

设 w 是当前读入的输入串;

q_s :在开始状态读入1时,有 $w=1$,进入状态 q_1 ;读入1时,有 $w=2$,进入状态 q_2 ;读入3时,有 $w=3$,进入状态 q_3 ;

q_0 :能引导自动机到达次状态的 w 除以4余数为0,因此, $w=4n+0$;

在此状态读入1,引导自动机到达下一状态的输入串为 $w1$,则 $w1=10(4n+0)+1=4\times 10n+1$,表明 $w1$ 属于 q_1 对应的等价类,所以自动机在 q_0 状态读入1,应该到达 q_1 状态;

在此状态读入2,引导自动机到达下一状态的输入串为 $w2$,则 $w2=10(4n+0)+2=4\times 10n+2$,表明 $w2$ 属于 q_2 对应的等价类,所以自动机在 q_0 状态读入2,应该到达 q_2 状态;

在此状态读入3,引导自动机到达下一状态的输入串为 $w3$,则 $w3=10(4n+0)+3=4\times 10n+3$,表明 $w3$ 属于 q_3 对应的等价类,所以自动机在 q_0 状态读入3,应该到达 q_3 状态;

q_1 :能引导自动机到达次状态的 w 除以4余数为1,因此, $w=4n+1$;

在此状态读入1,引导自动机到达下一状态的输入串为 $w1$,则 $w1=10(4n+1)+1=4\times 10n+11$,表明 $w1$ 属于 q_3 对应的等价类,所以自动机在 q_1 状态读入1,应该到达 q_3 状态;

在此状态读入2,引导自动机到达下一状态的输入串为 $w2$,则 $w2=10(4n+1)+2=4\times 10n+12$,表明 $w2$ 属于 q_0 对应的等价类,所以自动机在 q_1 状态读入2,应该到达 q_0 状态;

在此状态读入3,引导自动机到达下一状态的输入串为 $w3$,则 $w3=10(4n+1)+3=4\times 10n+13$,表明 $w3$ 属于 q_1 对应的等价类,所以自动机在 q_1 状态读入3,应该到达 q_1 状态;

q_2 :能引导自动机到达次状态的 w 除以4余数为2,因此, $w=4n+2$;

在此状态读入1,引导自动机到达下一状态的输入串为

$w1$,则 $w1=10(4n+2)+1=4\times 10n+21$,表明 $w1$ 属于 q_1 对应的等价类,所以自动机在 q_2 状态读入1,应该到达 q_1 状态;

在此状态读入2,引导自动机到达下一状态的输入串为 $w2$,则 $w2=10(4n+2)+2=4\times 10n+22$,表明 $w2$ 属于 q_2 对应的等价类,所以自动机在 q_2 状态读入2,应该到达 q_2 状态;

在此状态读入3,引导自动机到达下一状态的输入串为 $w3$,则 $w3=10(4n+2)+3=4\times 10n+23$,表明 $w3$ 属于 q_3 对应的等价类,所以自动机在 q_2 状态读入3,应该到达 q_3 状态;

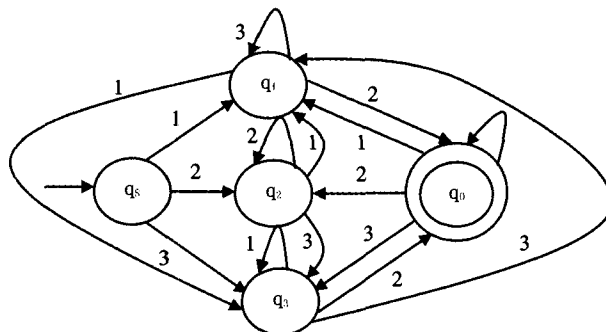
q_3 :能引导自动机到达次状态的 w 除以4余数为3,因此, $w=4n+3$;

在此状态读入1,引导自动机到达下一状态的输入串为 $w1$,则 $w1=10(4n+3)+1=4\times 10n+31$,表明 $w1$ 属于 q_3 对应的等价类,所以自动机在 q_3 状态读入1,应该到达 q_3 状态;

在此状态读入2,引导自动机到达下一状态的输入串为 $w2$,则 $w2=10(4n+3)+2=4\times 10n+32$,表明 $w2$ 属于 q_0 对应的等价类,所以自动机在 q_3 状态读入2,应该到达 q_0 状态;

在此状态读入3,引导自动机到达下一状态的输入串为 $w3$,则 $w3=10(4n+3)+3=4\times 10n+33$,表明 $w3$ 属于 q_1 对应的等价类,所以自动机在 q_3 状态读入3,应该到达 q_1 状态。

综上所述,得有限状态自动机为:



总结 本文提出了根据等价类构造一类有限状态自动机的方法。该方法可以针对所有 M 进制的数字串构成的语言,而且满足语言中的所有数字串能够整除任意正整数 N 。

参考文献

- 1 陈有祺. 形式语言与自动机[M]. 南京:南开大学出版社,1999
- 2 Sipser M 著. 张立昂,等译. 计算理论导论[M]. 北京:机械工业出版社,1996
- 3 蒋宗礼. 形式语言与自动机理论[M]. 北京:清华大学出版社,2003
- 4 阮传概. 近世代数及其应用[M]. 北京:邮电大学出版社,2002