形式语言与自动机期末速通

1. 绪论

1.1 集合

1.1.1 对称差

[**定义1.1.1**] 对集合A和集合B,称属于A但不属于B的元素、属于B但不属于A的元素组成的集合为A与B的**对称差**,记作$A \oplus B$,即$A \oplus B = \{a \mid (a \in A \wedge a \notin B) \vee (a \notin A \wedge a \in B)\}$.

[**定理1.1.1**] [**对称差的性质**] 对集合A和B,$A \oplus B = (A \bigcup B) - (A \bigcap B) = (A - B) \bigcup (B - A)$.

1.1.2 Cartesian积

[**定义1.1.2**] 集合A与集合B的**Cartesian积**是一个集合,该集合包含所有的有序数对(a, b),其中$a \in A, b \in B$.集合A与集合B的Cartesian积记作$A \times B$,即$A \times B = \{(a, b) \mid a \in A, b \in B\}$.

[**例1.1.1**] 设集合$A = \{1, 2, 3\}$,集合$B = \{r, y, b\}$,则它们的Cartesian积:

$A \times B = \{(1, r), (1, y), (1, b), (2, r), (2, y), (2, b), (3, r), (3, y), (3, b)\}$.

1.1.3 幂集

[**定义1.1.3**] 集合A的**幂集**是一个集合,该集合由A的所有子集组成,记作2^A,即$2^A = \{B \mid B \subseteq A\}$.

[**例1.1.2**] 集合$A = \{1, 2, 3\}$的幂集:

$2^A = \{\varnothing, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}$.

1.2 关系

1.2.1 二元关系

[**定义1.2.1**] 对集合A和集合B,任意的集合$R \subseteq A \times B$称为A到B的一个**二元关系**,其中A称为**定义域**,B称为**值域**,$(a, b) \in R$可表示为aRb.特别地,$A = B$时,称R是A上的二元关系.

[**例1.2.1**] 对集合$A = \{1, 3, 6\}$和集合$B = \{2, 5, 7\}$,"小于"的二元关系:

$R_< = \{(1, 2), (1, 5), (1, 7), (3, 5), (3, 7), (6, 7)\}$.

1.2.2 等价关系

[定义1.2.2] 二元关系的三岐性是指自反性、对称性、传递性,称具有三岐性的二元关系为**等价关系**.等价关系划分了集合,集合S上的等价关系R确定了S的一个等价分类.

具体地,对集合A上的二元关系R:

(1)自反性:若$a \in A$,则$(a, a) \in R$.

(2)对称性:若$a, b \in A$且$(a, b) \in R$,则$(b, a) \in R$.

(3)传递性:若$a, b, c \in A$,且$(a, b), (b, c) \in R$,则$(a, c) \in R$.

[定义1.2.3] 称对集合S划分$S_1, \cdots, S_n, \cdots$为S关于等价关系R的**等价划分**,如果满足下列性质,其中每个$S_i \ \ (i = 1, 2, \cdots)$称为**一个等价类**.

(1)$S = \bigcup\limits_{i=0}^{\infty} S_i$.

(2)$S_i \bigcap S_j = \varnothing \ \ (i \neq j)$.

(3)对任一等价类中的两元素a, b,有aRb恒成立.

(4)对不同等价类中的两元素a, b,有aRb恒不成立.

[定义1.2.4] 等价关系R将集合S划分成的等价类的个数称为R在S上的**指数**.若R将S划分为有穷多个等价类,则称R有**有穷指数**;若R将S划分为无穷多个等价类,则称R有**无穷指数**.

1.2.3 关系的合成

[定义1.2.5] 对集合A、集合B和集合C,设$R_1 \subseteq A \times B$是A到B的二元关系,$R_2 \subseteq B \times C$是B到C的二元关系,则二元关系R_1与二元关系R_2的**合成**R_1R_2是A到C的二元关系,即$R_1R_2 = \{(a, c) \mid \exists (a, b) \in R_1, (b, c) \in R_2\}$.

[例1.2.2]

(1)设R_1为"父子关系",R_2为"父女关系",则它们的合成R_1R_2为"祖孙女关系".

(2)设集合$R_1 = \{(1, a), (2, b), (3, c)\}$,集合$R_2 = \{(a, 4), (c, 5)\}$,则它们的合成:$R_1R_2 = \{(1, 4), (3, 5)\}$.

1.2.4 关系的闭包

[定义1.2.6] 设P是关于关系的性质的集合,关系R的P**闭包**是包含R且有P中所有性质的最小关系.

[定义1.2.7] 关系R的**正闭包**R^+满足如下三个性质:

(1)$R \subset R^+$.

(2)若$(a, b), (b, c) \in R^+$,则$(a, c) \in R^+$.

(3)除(1)(2)外,R^+不再包含其他元素.

[定义1.2.8] 有传递性的闭包称为**传递闭包**,如R^+.

[**定理1.2.1**] 对集合S上的二元关系R,有:

(1)$R^+ = R \bigcup R^2 \bigcup \cdots \bigcup R^n \bigcup \cdots$.

(2)S为有限集时,有$R^+ = R \bigcup R^2 \bigcup \cdots \bigcup R^{|S|}$.

[**定义1.2.9**] 关系R的**Kleene闭包**R^*满足如下三个性质:

(1)$R^0 \subseteq R^*, R \subseteq R^*$.

(2)若$(a,b),(b,c) \in R^*$,则$(a,c) \in R^*$.

(3)除(1)(2)外,R^+不再包含其他元素.

[**定义1.2.10**] 有自反性、传递性的闭包称为**自反传递闭包**,如R^*.

[**定理1.2.2**] 对集合S上的二元关系R,有:

(1)$R* = R^0 \bigcup R^+ = R^0 \bigcup R \bigcup R^2 \bigcup \cdots \bigcup R^n \bigcup \cdots$.

(2)S为有限集时,有$R^* = R^0 \bigcup R \bigcup R^2 \bigcup \cdots \bigcup R^{|S|}$.

[**例1.2.3**] 设$R_1 = \{(a,b),(c,d),(b,d),(b,b),(d,e)\}$和$R_2 = \{(a,a),(b,c),(d,c),(e,d),(c,a)\}$是集合$S = \{a,b,c,d,e\}$上的二元关系.

$R_1R_2 = \{(a,c),(c,c),(b,c),(d,d)\}$.

$R_2R_1 = \{(a,b),(c,b),(b,d),(d,d),(e,e)\}$.

$R_1^+ = \{(a,b),(c,d),(b,d),(b,b),(d,e),(a,d),(c,e),(b,e),(a,e)\}$,即抄一遍$R_1$后补上由传递性生成的元素.

$R_1^* = \{(a,b),(c,d),(b,d),(b,b),(d,e),(a,d),(c,e),(b,e),(a,e),(a,a),(c,c),(d,d),(e,e)\}$,

即抄一遍R_1^+后补上由自反性生成的元素.

1.3 递归定义与归纳证明

[**例1.3.1**] 对集合S上的二元关系R,定义R的n次幂R^n为:

(1)$R^0 = \{(a,a) \mid a \in S\}$.

(2)$R^i = R^{i-1}R\ \ (i \geq 1)$.

[**注**] 注意(2)中等式右边R^{i-1}在前.

[**例1.3.2**] 求证:对有穷集A,有$\left|2^A\right| = 2^{|A|}$.

[**证**]

(1)基础:$|A| = 0$,即$A = \varnothing$时,有$\left|2^A\right| = |\{\varnothing\}| = 1$.

(2)归纳:设$|A| = n\ \ (n \geq 0)$时结论成立,下证$|A| = n+1$时结论成立.

设$|A|=n$时的幂集为2^B,$|A|=n+1$时新加入的元素为a,

则$2^A=2^B\bigcup\{C\bigcup\{a\}\mid C\in 2^B\}$,其中$2^B\bigcap\{C\bigcup\{a\}\mid C\in 2^B\}=\varnothing$.

故$\left|2^A\right|=\left|2^B\right|+\left|\{C\bigcup\{a\}\mid C\in 2^B\}\right|=\left|2^B\right|+\left|2^B\right|=2\cdot\left|2^B\right|=2\cdot 2^{|B|}=2^{|B|+1}=2^{|A|}$.

(3)由归纳法原理:结论对任意有穷集成立.

1.4 语言

[**定义1.4.1**] 一个语言的**字母表**是一个非空有穷集,记作Σ,其中的元素称为字母表的一个**字母**或**符号**或**字符**.字符由两个特性:①整体性,又称不可分割性;②可辨认性,又称可区分性.

[**注**] 一般用小写字母中较靠前的字母表示字母表中的字母,如$a,b,c,\cdots$.

[**例1.4.1**]

(1)$\varnothing$不是字母表,因为不满足非空性.

(2)集合$\{a,b,a,c\}$不是字母表,因为不满足可区分性.

(3)集合$\{0,1,2,\cdots,n,\cdots\}$不是字母表,因为不满足有穷性.

[**定义1.4.2**] 对两个字母表Σ_1和Σ_2,定义它们的**乘积**$\Sigma_1\Sigma_2=\{ab\mid a\in\Sigma_1,b\in\Sigma_2\}$.

[**例1.4.2**]

(1)$\{0,1\}\{0,1\}=\{00,01,10,11\}$.

(2)$\{0,1\}\{a,b,c,d\}=\{0a,0b,0c,0d,1a,1b,1c,1d\}$.

(3)$\{aa,ab,bb\}\{0,1\}=\{aa0,aa1,ab0,ab1,bb0,bb1\}$.

[**定义1.4.3**] 对字母表Σ,定义其n次幂:

(1)$\Sigma^0=\{\varepsilon\}$,其中ε由Σ中的0个字符组成,即空串.

(2)$\Sigma^n=\Sigma^{n-1}\Sigma$.

[**注**] 注意(2)中等式右边Σ^{n-1}在前.

[**定义1.4.4**] 对字母表Σ,定义其**正闭包**$\Sigma^+=\{x\mid x\text{是}\Sigma\text{中的至少一个字符连接而成的字符串}\}$,定义其**Kleene闭包**$\Sigma^*=\{x\mid x\text{是}\Sigma\text{中的若干个(含0个)字符连接而成的字符串}\}$.

[**定理1.4.1**] 对字母表Σ,

(1)其正闭包$\Sigma^+=\Sigma\bigcup\Sigma^2\bigcup\cdots\bigcup\Sigma^n\bigcup\cdots$.

(2)其Kleene闭包$\Sigma^*=\Sigma^0\bigcup\Sigma^+=\Sigma^0\bigcup\Sigma\bigcup\Sigma^2\bigcup\cdots\bigcup\Sigma^n\bigcup\cdots$.

[**例1.4.3**]

(1)$\{0,1\}^+=\{0,1,00,01,10,11,000,001,010,\cdots\}$,即所有非空二进制串.

$\{0,1\}^*=\{\varepsilon,0,1,00,01,10,11,000,001,010,\cdots\}$,即所有二进制串(含空串).

(2)$\{a,b,c,d\}^+=\{a,b,c,d,aa,ab,ac,ad,ba,bb,bc,bd,\cdots\}$.

$\{a,b,c,d\}^*=\{\varepsilon,a,b,c,d,aa,ab,ac,ad,ba,bb,bc,bd,\cdots\}$.

[**定义1.4.5**] 对字母表Σ,称$\forall x\in\Sigma^*$为Σ上的一个**句子**,又称**字**、**(字符、符号)行**、**(字符、符号)串**.两句子**相等**当且仅当它们对应位置上的字符都相等.

[**注**] 一般用小写字母中较靠后的字母表示字母表上的句子,如$x,y,z,\cdots$.

[**定义1.4.6**] 对句子$x,y\in\Sigma^*$,句子x,y的**并置**(又称**连结**)是由串x直接相接串y所组成的字符串,记作xy.

[**定义1.4.7**] 对句子$x,y\in\Sigma^*$和字母$a\in\Sigma$,句子xay中的a称为a在该句子中的**一个出现**.特别地,$x=\varepsilon$时,a的出现称为该字符串的**首字母**;$y=\varepsilon$时,a的出现称为该字符串的**尾字母**.对句子$x\in\Sigma^*$,称其中字母出现的总个数为该句子的**长度**,记作$|x|$.特别地,长度为0的句子称为**空句子**,记作ε.

[**注**] $\{\varepsilon\}\neq\varnothing$.

[**定义1.4.8**] 定义句子$x\in\Sigma^*$的n次幂:

(1)$x^0=\varepsilon$.

(2)$x^n=x^{n-1}x$.

[**例1.4.4**] 对句子$x=001$,

(1)$x^0=\varepsilon$.

(2)$x^4=001001001001$.

[**定义1.4.8**] 设句子$x,y,z\in\Sigma^*$,其中$x=yz$.

(1)称y为x的**前缀**.特别地,若$z\neq\varepsilon$,则称y为x的**真前缀**.

(2)称z为x的**后缀**.特别地,若$y\neq\varepsilon$,则称z为x的**真后缀**.

[**例1.4.5**] 对句子$abaabb\in\{a,b\}^*$,

(1)前缀:$\varepsilon,a,ab,aba,abaa,abaab,abaabb$.

真前缀:$a,ab,aba,abaa,abaab,abaabb$.

(2)后缀:$\varepsilon,b,bb,abb,aabb,baabb,abaabb$.

真后缀:$b,bb,abb,aabb,baabb,abaabb$.

[**定义1.4.9**] 设句子$w, x, y, z \in \Sigma^*$.若$w = xyz$,则称y为w的**子串**.

[**定义1.4.10**] 对字母表Σ,称$L \subset \Sigma^*$为Σ上的一个**语言**,称$x \in L$为L中的一个**句子**.

[**定义1.4.11**] 对语言$L_1 \subseteq \Sigma_1^*$和语言$L_2 \subseteq \Sigma_2^*$,定义它们的**乘积**为语言$L_1L_2 = \{xy \mid x \in L_1, y \in L_2\}$,它是字母表$\Sigma_1 \bigcup \Sigma_2$上的语言.

[**定义1.4.12**] 对语言$L \in \Sigma^*$,定义L的n次幂:

(1)$n = 0$时,$L^0 = \{\varepsilon\}$.

(4)$n \geq 1$时,$L^n = L^{n-1}L$.

[**注**] 注意(2)中等式右边L^{n-1}在前.

[**定义1.4.13**] 对语言$L \in \Sigma^*$,定义:

(1)其正闭包$L^+ = L \bigcup L^2 \bigcup \cdots \bigcup L^n \bigcup \cdots$.

(4)其Kleene闭包$L^* = L^0 \bigcup L^+ = L^0 \bigcup L \bigcup L^2 \bigcup \cdots \bigcup L^n \bigcup \cdots$.

[**例1.4.6**] 下列集合都是字母表$\{0, 1\}$上的语言:

(1)$\{0, 1\}$.

(2)$\{0, 1, 00, 11\}$.

(3)$\{00, 11\}^*$.

(4)$\{0\}\{0, 1\}^*\{1\}$.

[**例1.4.7**] 考察下列字母表$\Sigma = \{0, 1\}$上的语言:

(1)$L_1 = \{0, 1\}$.

(2)$L_2 = \{00, 01, 10, 11\}$.

(3)$L_3 = \{0, 1, 00, 01, 10, 11, 000, \cdots\} = \Sigma^+$.

(4)$L_4 = \{\varepsilon, 0, 1, 00, 01, 10, 11, 000, \cdots\} = \Sigma^*$.

(5)$L_5 = \{0^n \mid n \geq 1\}$.

(6)$L_6 = \{0^n 1^n \mid n \geq 1\}$.

(7)$L_7 = \{1^n \mid n \geq 1\}$.

(8)$L_8 = \{0^n 1^m \mid n, m \geq 1\}$.

(9)$L_9 = \{0^n 1^n 0^n \mid n \geq 1\}$.

(10)$L_{10} = \{0^n 1^m 0^k \mid n, m, k \geq 1\}$.

(11)$L_{11} = \{x \mid x \in \Sigma^+ \text{且} x \text{中} 0 \text{和} 1 \text{的个数相同}\}$.

[**注**] ①上述语言都是L_4的子集,也称**子语言**.

②L_1, L_2都是有穷语言,其他的语言都是无穷语言.

③L_1是Σ上所有长度为1的句子组成的语言.

④L_2是Σ上所有长度为2的句子组成的语言.

⑤$L_5 L_7 \neq L_6$,但$L_5 L_7 = L_8$,即不同语言中的n可能不同.

⑥$L_6 \subseteq L_5 L_7, L_9 \subseteq L_{10}$.

⑦L_{11}不要求所有0在所有1之前,故$L_6 \subset L_{11}$.

[**例1.4.8**] 设字母表$\Sigma = \{0, 1\}$,写出下列语言的形式表示.

(1)所有以0开头的串:$\{0\}\{0,1\}^*$.

(2)所有以11开头、以11结尾的串:$\{11\}\{0,1\}^*\{11\} \bigcup \{11, 111\}$.

注意写出语言的乘积的表示后需检查是否有遗漏的串.

(3)所有所有包含子串001的串:$\{0,1\}^*\{001\}\{0,1\}^*$.

(4)所有正数第10个字符是0的串:$\{0,1\}^9\{0\}\{0,1\}^*$.

(5)所有不包含连续3个0的串:$\{1, 01, 001\}^* \bigcup \{10, 100\}$.