第一章 经典集合论

郑征

zhengz@buaa. edu. cn

北京航空航天大学

软件与控制研究室

2018年9月11日星期二

第1讲 集合的概念与运算

1. 集合的概念

2. 集合之间的关系

3. 集合的运算

4. 文氏图、容斥原理

1. 集合的概念

集合论(set theory)

- 十九世纪数学最伟大成就之一
- 集合论体系
  - 朴素(naive)集合论
  - 公理(axiomatic)集合论
- 创始人康托(Cantor)

Georg Ferdinand Philip Cantor
1845 ~ 1918

德国数学家, 集合论创始人.

集合论

- 集合语言是现代数学的基本语言，使用集合语言可以简洁准确的表达数学中的一些内容。
  - 数学里有自然语言、符号语言、图形语言，还有图表语言等，集合就是一种特殊的符号语言。
- 集合作为一个数学概念，对于数学中的分类思想，可以起到一个促进的作用。集合主要是要把各种不同的事物能按类刻划清楚。古语有云：人以类聚，物以群分。这实际上就是一种集合思想。

- 一个数据库由若干个数据表构成
  - 每个数据表可以认为是若干记录的集合
- C语言中的整型数组是若干整数的集合
  - 如果不考虑次序关系
- 一个班是若干学生的集合
- 一个公司是若干员工的集合

将具有某种共同特征的事物汇集到一起组成一个整体

什么是集合(set)

- 集合：将具有某种共同特征的事物汇集到一起组成一个整体，这个整体就称为集合。
  - 组成集合的事物叫做该集合的元素；
  - 用大写英文字母A,B,C,…表示集合；用小写英文字母a,b,c,…表示元素。

- $a \in A$：表示a是A的元素，读作“a属于A”
  $a \notin A$：表示a不是A的元素，读作“a不属于A”

朴素集合论中元素和集合之间有且仅有两种关系

集合的表示

- 列举法
- 描述法
- 特征函数法

规范的符号定义和表示是非常重要的，不仅仅在数学里，在程序设计乃至科技论文写作中都是如此。

Why?

统一性；无二义性。。。1$\in$N, 1.5 $\notin$N

列举法(roster)

- 列出集合中的全体元素，元素之间用逗号分开，然后用花括号括起来，例如

$$A=\{a,b,c,d,\ldots,x,y,z\}$$

$$B=\{0,1,2,3,4,5,6,7,8,9\}$$

太多了，太长了，太烦了

- 集合中的元素不规定顺序

$$C=\{2,1\}=\{1,2\}$$

- 集合中的元素各不相同(多重集除外)

$$C=\{2,1,1,2\}=\{2,1\}$$

多重集(multiple set)

- 多重集：允许元素多次重复出现的集合
- 元素的重复度: 元素的出现次数(≥0).
- 例如：设A={a,a,b,b,c}是多重集

    元素a,b的重复度是2

    元素c的重复度是1

    元素d的重复度是0

描述法(defining predicate)

- 用谓词P(x)表示x具有性质P， 用{x|P(x)}表示具有性质 P 的集合，例如
- P_1 (x): x是英文字母

  A={x|P_1 (x)}={x| x是英文字母}

  ={a,b,c,d,…,x,y,z}
- P_2 (x): x是十进制数字

  B={x|P_2(x)}= {x|x是十进制数字}

  ={0,1,2,3,4,5,6,7,8,9}

谓词：用于表示集合元素（共同）特性的词语。

- 请举例：
  - 哪些集合只能用谓词描述法，不能用列举法。
  - 实数集，整数集。。。
- Question 1
  - A=(1, 2, 3);
  - A={1, 2, 3}.
  - Which one is right?
- Question 3
  - A={1、 2、 3}. Right?

特征函数法(characteristic function)

- 集合A的特征函数是$\chi_A(x)$:

$$\chi_A(x) = \begin{cases} 1，若 x \in A \\ 0，若 x \notin A \end{cases}$$

- 对多重集, $\chi_A(x)$=x在A中的重复度

模糊集

集合元素的特性

- 特性１：无序性
  - 集合中的元素是无序的；
    - {张三，李四，王二}={李四，王二，张三}
    - {a, b, c}={c, b, a}
- 特性２：集合中的元素可以是任何类型的事物，集合的元素也可以是集合。
  - 递归定义；C 语言。
- 特性３：如果不考虑多重集，集合中的元素彼此不同，重复出现的元素自动删除。
  - {1, 2, 3, 4, 1, 2, 5}={1, 2, 3, 4, 5}

研究热点

模糊集

粗糙集

不确定性

从精确到模糊

- 精确
  - 答案确定：要么是，要么不是
  - $f : A \to \{0,1\}$
  - 他是学生？他不是学生？

- 模糊
  - 答案不定：也许是，也许不是，也许介于之间
  - $\mu_A : U \to [0,1]$
  - 他是成年人？他不是成年人？他大概是成年人？

模糊集

在普通集合中，论域中的元素（如a）与集合（如A）之间的关系是属于（a∈A），或者不属于（a∉A），它所描述的是非此即彼的清晰概念。但在现实生活中并不是所有的事物都能用清晰的概念来描述，如:

风的强弱

人的胖瘦

年龄大小

个子高低

2. 集合之间的关系

子集、相等、真子集

空集、全集

幂集、n元集、有限集

子集(subset) 和相等（equal）

- B是A的子集：B包含于A, A包含B:

等价于　　　　蕴含：则，肯定，保证

$$B \subseteq A \Leftrightarrow \forall x(x \in B \rightarrow x \in A)$$

数理逻辑

如果B中的每个元素都是A 中的元素。

- B不是A的子集:

$$B \not\subseteq A$$

- 相等:　$A=B \Leftrightarrow A \subseteq B \wedge B \subseteq A$

空集

- 空集:不含任何元素的集合称为空集,记作$\varnothing$
- 例如, $\{x \in R | x^2 + 1 = 0\}$
- 定理1: 对任意集合A, $\varnothing \subseteq A$。
- 推论: 空集是唯一的.

证明: 设$\varnothing_1$与$\varnothing_2$都是空集, 则

$$\varnothing_1 \subseteq \varnothing_2 \wedge \varnothing_2 \subseteq \varnothing_1 \Leftrightarrow \varnothing_1 = \varnothing_2 . \quad \#$$

全集

- 全集: 如果限定所讨论的集合都是某个集合的子集,则称这个集合是全集,记作E.
- 全集是相对的, 视情况而定, 因此不唯一.例如, 讨论(a,b)区间里的实数性质时, 可以选 $E=(a,b)$, $E=[a,b)$, $E=(a,b]$, $E=[a,b]$, $E=(a,+\infty)$,$E=(-\infty,+\infty)$等

幂集(power set)

- 幂集: A的全体子集组成的集合,称为A的幂集,记作P(A)

$$P(A)=\{x|x\subseteq A\}$$

- 注意: $x\in P(A) \Leftrightarrow x\subseteq A$
- 例子: $A=\{a,b\}$, $P(A)=\{\varnothing,\{a\},\{b\},\{a,b\}\}$.
- Question: $P(\varnothing)=?$

n元集(n-set)

- n元集: 含有n个元素的集合称为n元集.
- 0元集: $\varnothing$
- 1元集(或单元集),如{a}, {b}, {$\varnothing$}, {{$\varnothing$}},…
- |A|: 表示集合A中的元素个数,

$$A\text{是}n\text{元集} \Leftrightarrow |A|=n$$

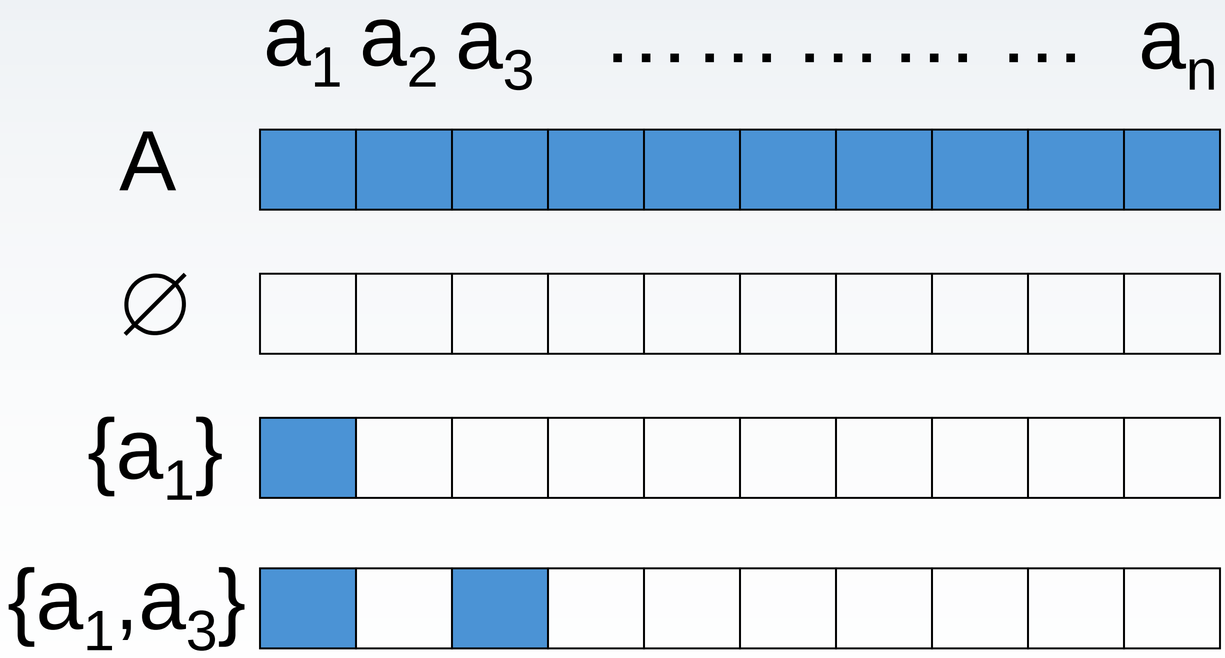
- 有限集 (fimite set): |A|是有限数, $|A|<\infty$, 也叫有穷集

在C++程序中，n元数组可以表示一个n元集
e.g. R=new int [10];C中用指针或者malloc

幂集(续)

- 定理: $|A|=n \Rightarrow |P(A)|=2^n$.

  证明: 每个子集对应一种染色,一共有2^n种不同染色. #



3.　集合之间的运算

- 并集
- 交集
- 相对补集
- 对称差
- 绝对补

并集(union)与交集(intersection)

- 并集:　所有属于A或属于B的元素组成的集合。

$$A \cup B = \{\, x \mid (x \in A) \overset{\text{或}}{\vee} (x \in B) \,\}$$

$$x \in A \cup B \Leftrightarrow (x \in A) \vee (x \in B)$$

分别从宏观角度和微观角度来看

初级并:

$$\bigcup_{i=1}^{n} A_i = A_1 \cup A_2 \cup \cdots \cup A_n$$

- 交集:　即属于A又属于B的元素组成的集合。

$$A \cap B = \{\, x \mid (x \in A) \wedge (x \in B) \,\}$$

$$x \in A \cap B \Leftrightarrow (x \in A) \wedge (x \in B)$$

初级交:

$$\bigcap_{i=1}^{n} A_i = A_1 \cap A_2 \cap \cdots \cap A_n$$

应用案例

- 软件测试中测试用例集合的选择
  - 测试用例1的运行可以覆盖的语句集合为S1;
  - 测试用例2的运行可以覆盖的语句集合为S2;
  - 。。。
  - 测试用例n的运行可以覆盖的语句集合为Sn;
  - 一个测试用例集合满足语句覆盖，即:
    - $S1 \cup S2 \ldots \cup Sn = E$，E为所有的语句。
  - 类似的还有条件覆盖，判定覆盖。。。

不相交(disjoint)

- 不相交： $A\cap B=\varnothing$

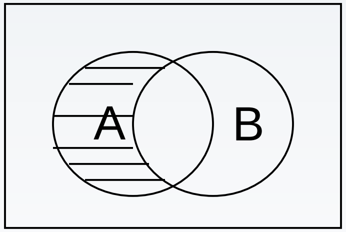
- 互不相交：设$A_1,A_2,\ldots$是可数多个集合, 若对于任意的$i\neq j$, 都有$A_i\cap A_j=\varnothing$, 则说它们互不相交

- 例: 设 $A_n=\{x\in R|n-1<x<n\}$, $n=1,2,\ldots,10$, 则$A_1,A_2,\ldots$是不相交的

相对补集(set difference)

- 相对补集（差集）：属于A而不属于B的全体元素,称为B对A的相对补集, 记作A-B

$$A\text{-}B = \{\, x \mid (x \in A) \wedge (x \notin B) \,\}$$



A-B

例：设A={1, 2, 3}, B={1, 3}, C={1, 2, 3, 4}，则

A-B= {1, 2, 3}-{1, 3}={2};

A-C={1, 2, 3}- {1, 2, 3, 4}=∅

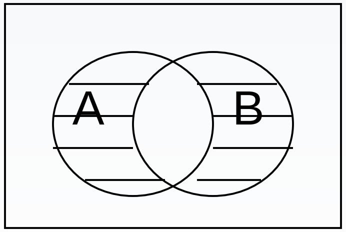
- A为失败的测试用例经过的语句集合；
- B为成功的测试用例经过的语句集合。

对称差(symmetric difference)

- 对称差: 属于A而不属于B, 或属于B而不属于A的全体元素（$(A-B)\cup(B-A)$）, 称为A与B的对称差, 记作$A\oplus B$

$$A\oplus B=\{x|(x\in A\wedge x\notin B)\vee(x\notin A\wedge x\in B)\}$$

- $A\oplus B=(A-B)\cup(B-A)=(A\cup B)-(A\cap B)$



$A\oplus B$

- A为失败的测试用例经过的语句;
- B为成功的测试用例经过的语句。
- A和B的对称差在缺陷定位中是最重要的语句。

绝对补(complement)

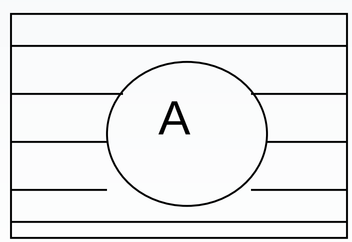
- 绝对补（补集）：由全集E中所有不属于A的元素组成的集合。

$$\sim A = E - A,\ E\text{是全集},\ A \subseteq E$$

A的补集等于全集E和A的相对补集（差集）。

$$\sim A = \{x \mid (x \in E \wedge x \notin A)\}$$

$$\sim A = \{x \in E \mid x \notin A)\}$$



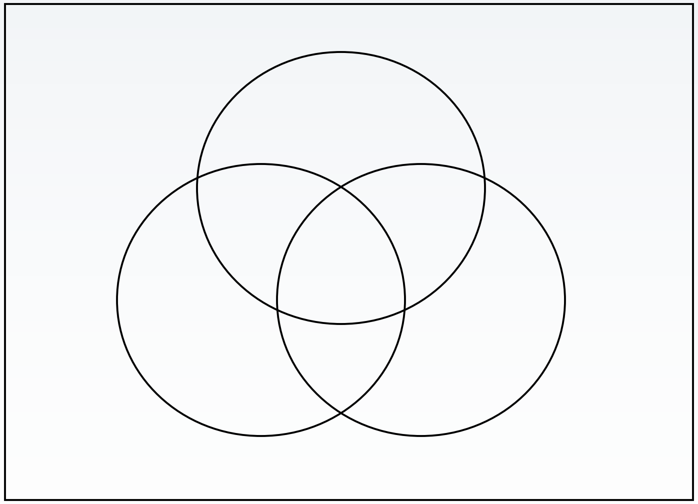
~A

注意：补集是相对于全集而言的，不是独立存在的。全集不同，补集也不同。

4. 文氏图、容斥原理

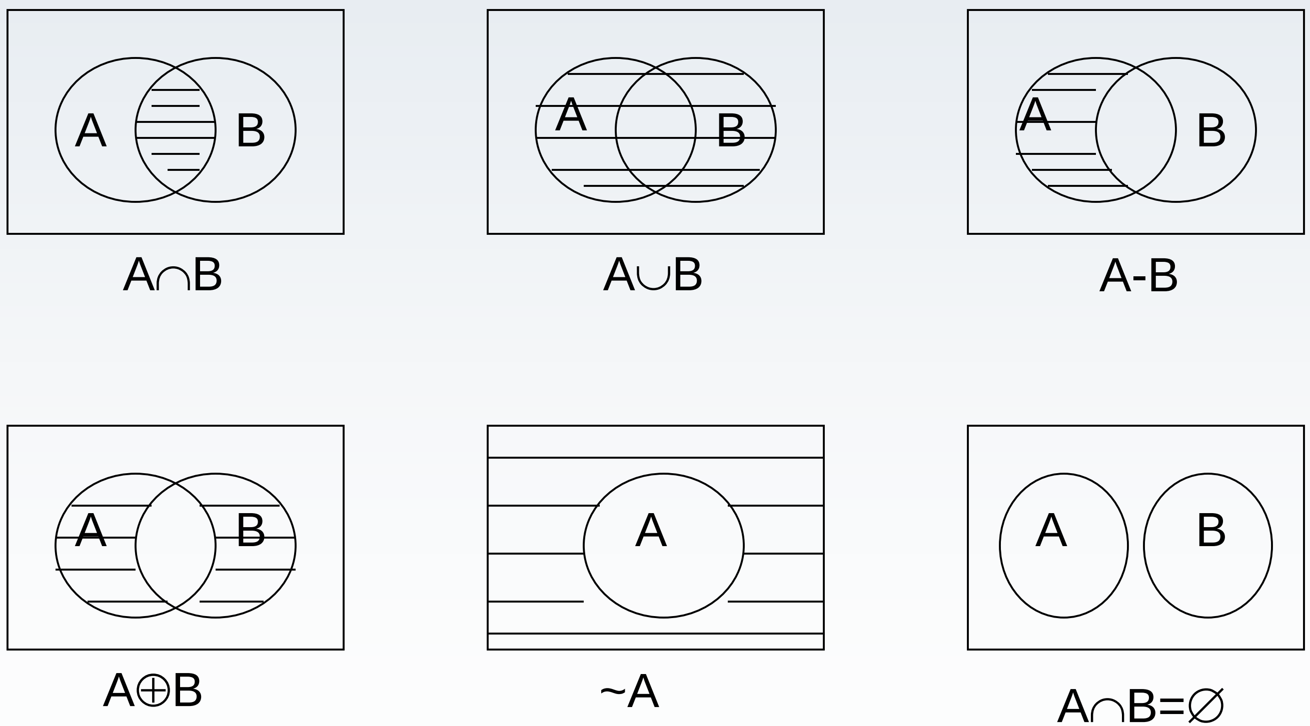
文氏图(Venn diagram)

- 文氏图: 平面上的n个圆(或椭圆),使得任何可能的相交部分, 都是非空的和连通的
- John Venn, 1834~1923
- 例:



文氏图(应用)

- 文氏图可表示集合运算(结果用阴影表示)



优点：形象、易于理解
人对图形的感知能力比文字（符号）强大

容斥原理(principle of inclusion/exclusion)

- 容斥原理(或包含排斥原理)

$$|\bigcup_{i=1}^{n} A_i| = \sum_{i=1}^{n} |A_i| - \sum_{i<j} |A_i \cap A_j|$$

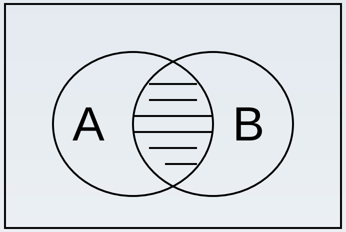
$$+ \sum_{i<j<k} |A_i \cap A_j \cap A_k| - \cdots + (-1)^{n-1} |A_1 \cap A_2 \cap \cdots \cap A_n|$$

数学归纳法进行证明

容斥原理(证明)

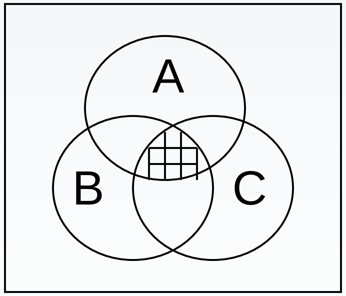
- n=2时的情况:

$$|A\cup B|=|A|+|B|-|A\cap B|$$



- 归纳证明: 以n=3为例:

$$|A\cup B\cup C| = |(A\cup B)\cup C| = |A\cup B|+|C|-|(A\cup B)\cap C|$$
$$= |A|+|B|-|A\cap B|+|C|-|(A\cap C)\cup(B\cap C)|$$
$$= |A|+|B|-|A\cap B|+|C|$$
$$-(|A\cap C|+|B\cap C|-|(A\cap C)\cap(B\cap C)|)$$
$$= |A|+|B|+|C|-|A\cap B|-|A\cap C|-|B\cap C|$$
$$+|A\cap B\cap C|$$



总结

- 集合概念: $\in$, $\varnothing$, E, $\subseteq$, $\subset$,
- 集合运算: $\cap$, $\cup$, -, $\oplus$, ~, P( )
- 文氏图
- 容斥原理

为什么离散数学中要研究集合论

- 集合：最基本最简单的离散数据结构
  - 计算机所“理解”知识和“数据”的重要组织方式；
  - 命题逻辑：若干规则的集合；
  - 机器学习：对样本集合的学习；
  - 图论：点的集合，边的集合；
  - 关系：二元组或n元组的集合；
  - 软件：代码的集合，函数的集合，子系统的集合，测试用例的集合。

为什么离散数学中要研究集合论

- 集合看似简单，但
  - 集合的抽象：
    - 是否可以从当前集合中总结出它内在的东西；*知识工程，聚类*
  - 集合间的关系：
    - 集合之间的关系可以刻画么？*学生集合，课程集合*
    - 可以根据一个集合来评价另一个集合么？*学习*
    - 不同类型集合之间的关系：*图（边集，点集）；*
  - 集合中的“特殊点”
    - *测试用例选择；软件缺陷代码选择*