第 5 章

函　数

函数是最基本的数学概念之一，也是一种最重要的数学工具，狭义上的函数 $y=f(x)$ 通常是在实数集合上进行的讨论。但广义的函数，实际是指一个量随另一个（或多个）量而变化的关系，因此，函数可以看作是一种特殊的关系。

例如，计算机中把输入、输出间的关系看成是一种函数；类似地，在开关理论、自动机理论和可计算性理论等领域中，函数也有着极其广泛的应用。

5.1　函数的概念

定义 5.1 函数　设 X、Y 为任意两个集合，如果 f 为 X 到 Y 的关系（$f\subseteq X\times Y$），且对每一个 $x\in X$，都有唯一的 $y\in Y$，使得 $<x,y>\in f$，则称 f 是 X 到 Y 的函数，记为 $f:X\rightarrow Y$ 或 $y=f(x)$。并称 x 为函数的自变元（或源点），y 为在函数 f 作用下 x 的像点（或函数值）。

在函数的定义中，集合 X 称为函数的定义域，记为 $\mathrm{dom}\ f=X$。而所有像点构成的集合称为函数 f 的值域或函数 f 的像，记为 $\mathrm{ran}\ f$ 或 $f(X)$，$\mathrm{ran}\ f\subseteq Y$。

根据函数的定义可知，函数是一种特殊的关系，其与一般性关系的区别有两点：

① 函数的定义域是 X，而不是 X 的真子集。即每个 $x\in X$ 都有像 $y\in Y$ 存在（如存在性）。

② 一个 x 只能对应唯一的一个 y（如唯一性）。

这两点区别也称为函数的两个要素。

函数定义的数学形式为：

$$f=\{<x,y>\mid x\in X\wedge y\in Y\wedge f(x)=y\}\text{ 或 }y=f(x)$$

例题 5.1　设 $X=\{1,5,p,\text{张明}\}$，$Y=\{2,q,7,9,G\}$，判断下列关系是否为函数，并求其定义域和值域。

$$f=\{<1,2>,<5,q>,<p,7>,<\text{张明},G>\}$$

【解】　从题目给出的关系可以看出，每一个 $x\in X$，都有唯一的 $y\in Y$，使得 $<x,y>\in f$，即 $f(1)=2$、$f(5)=q$、$f(p)=7$、$f(\text{张明})=G$，因此，关系 f 是函数。

$\mathrm{dom}\ f=X$

$\mathrm{ran}\ f=\{2,q,7,G\}$

例题 5.2　判断下列关系中哪些能构成函数，并说明理由。

① $f=\{<x_1,x_2>\mid x_1、x_2\in N，\text{且 } x_1+x_2<10\}$；

② $f=\{<y_1,y_2>|y_1、y_2\in R$，且 $y_2^2=y_1\}$；

③ $f=\{<x_1,x_2>|x_1、x_2\in N,x_2$ 为小于 x_1 的素数个数$\}$。

【解】 ① 不能构成函数。因为 x_1 不能取定义域中所有的值，且一个 x_1 可能对应多个 x_2。

② 不能构成函数。因为一个 y_1 对应两个 y_2，故也不是函数。

③ 能够成为函数。

从函数的定义可以知道，$X\times Y$ 的子集并不能都成为 X 到 Y 的函数。

例题 5.3 设集合 $X=\{a,b,c\}$、$Y=\{0,1\}$，试写出所有从 X 到 Y 的函数。

【解】 $X\times Y=\{<a,0>,<b,0>,<c,0>,<a,1>,<b,1>,<c,1>\}$，$X\times Y$ 有 $2^6=64$ 个可能的子集，但其中只有 $2^3=8$ 个子集可定义为从 X 到 Y 的函数：

$f_0=\{<a,0>,<b,0>,<c,0>\}$　　$f_1=\{<a,0>,<b,0>,<c,1>\}$

$f_2=\{<a,0>,<b,1>,<c,0>\}$　　$f_3=\{<a,0>,<b,1>,<c,1>\}$

$f_4=\{<a,1>,<b,0>,<c,0>\}$　　$f_5=\{<a,1>,<b,0>,<c,1>\}$

$f_6=\{<a,1>,<b,1>,<c,0>\}$　　$f_7=\{<a,1>,<b,1>,<c,1>\}$

设 X 和 Y 都为有限集，分别有 m 个和 n 个不同元素，由于从 X 到 Y 任意一个函数的定义域都是 X，所以这些函数中每一个都有 m 个序偶。另外任何元素 $x\in X$，可以有 Y 的 n 个元素中任何一个作为它的像，故共有 n^m 个不同的函数。在例题 5.3 中，$n=2$、$m=3$，故应有 2^3 个不同的函数。

今后我们用符号 Y^X 表示从 X 到 Y 的所有函数的集合，甚至当 X 和 Y 是无限集时，也可用这个符号。

因为函数是序偶的集合，所以两个函数相等可借用集合相等的概念予以定义。

定义 5.2 函数相等 设函数 $f:A\rightarrow B$，$g:C\rightarrow D$，如果 $A=C$、$B=D$，且对所有 $x\in A$ 和 $x\in C$，都有 $f(x)=g(x)$，则称函数 f 等于函数 g，记为 $f=g$。

如果 $A\subseteq C$、$B=D$，且对每一个 $x\in A$，都有 $f(x)=g(x)$，则称函数 f 包含于函数 g，记为 $f\subseteq g$。

5.2 几种特殊的函数

函数描述了集合 X 中元素和集合 Y 中元素之间的特殊对应关系，这种对应关系体现了 X 中元素在 Y 上的投射性质，因此，函数也可以称为映射。映射关系可以是一对一的，也可以是多对一的。同时，函数的值域可以是集合 Y 的一个真子集，也可以是集合 Y 的全部。这些不同的情况，形成了一些特殊的函数(入射、满射、双射、恒等、常函数等)。

定义 5.3 入射函数 设 f 是集合 X 到 Y 的函数，如果 X 中任何 2 个元素都没有相同的像，则称 f 是集合 X 到 Y 的入射(单射、一对一映射)函数。

如果 $f:X\rightarrow Y$ 是入射函数，则对于任意的 $x_1、x_2\in X$，有：

$$x_1\neq x_2\Rightarrow f(x_1)\neq f(x_2)$$

或者

$$f(x_1)=f(x_2)\Rightarrow x_1=x_2$$

入射函数也称单射函数或者一对一的函数。

例如，函数 $f: \{a,b\} \to \{2,4,6\}$ 为 $f(a)=2$、$f(b)=6$，则这个函数是入射函数。

定义 5.4 满射函数　设 f 是集合 X 到 Y 的函数，如果 $\operatorname{ran} f=Y$，即 Y 的每一个元素都是 X 中一个或者多个元素的像点，则称 f 为满射函数(或 X 到 Y 上的映射)。

如果 $f:X\to Y$ 是满射，则对与任意 $y\in Y$，必存在 $x\in X$，使 $f(x)=y$ 成立。

例如，$A=\{a,b,c,d\}$，$B=\{1,2,3\}$，如果 $f:A\to B$ 为 $f(a)=1$、$f(b)=1$、$f(c)=3$、$f(d)=2$。则 f 是满射的。

定义 5.5 双射函数　设 f 是集合 X 到 Y 的函数，如果 f 既是 X 到 Y 的入射函数，又是 X 到 Y 的满射函数，则称 f 为 X 到 Y 的双射函数。双射函数也称一一对应的函数，Y 中的每一元素在 X 中都有且仅有一个源点。

例如，令 $[a,b]$ 表示实数的闭区间，即 $[a,b]=\{x\mid a\leqslant x\leqslant b\}$。令 $f:[0,1]\to[a,b]$，且 $f(x)=(b-a)x+a$。这个函数就是双射函数。

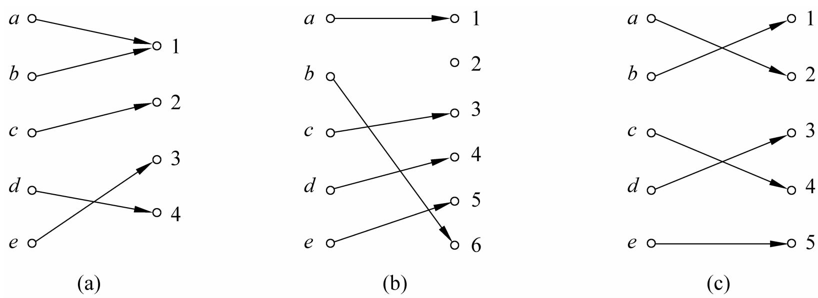

图 5.1　几种关系表示的函数

例题 5.4　判断图 5.1(a)、图 5.1(b)和图 5.1(c)所示关系图表示的函数是哪一种函数。

【解】　从关系图可以看出，图 5.1(a)和图 5.1(c)是满射函数；图 5.1(b)和图 5.1(c)是入射函数；而图 5.1(c)是双射函数。

定理 5.1　令 X 和 Y 为有限集，若 X 和 Y 的元素个数相同，即 $|X|=|Y|$，则 $f:X\to Y$ 是入射的，当且仅当它是一个满射。

【证明】　必要性。

若 f 是入射的，则 $|X|=|f(X)|$(一对一映射源点的个数=像的个数)，所以 $|f(X)|=|Y|$。而从函数 f 的定义可知 $f(X)\subseteq Y$，又因为 Y 是有限集合，元素个数固定，所以 Y 的每个元素都是 X 中元素的像点，即 f 是满射的。

充分性。

若 f 是满射的，则根据满射的定义可知 $f(X)=Y$，所以 $|X|=|Y|=|f(X)|$，又因为 X 是有限集合，所以 X 中不存在 2 个元素有相同的像点，所以 $f:X\to Y$ 是入射的。

本定理证毕。

注意：因为无限集合无法计算元素个数，所以本定理不适用于无限集合上的映射。

定义 5.6 常函数　设函数 $f:X\to Y$，如果存在某个 $y_0\in Y$，对于每个 $x\in X$ 都有 $f(x)=y_0$，即 $f(X)=\{y_0\}$，则称函数 f 为常函数。

定义 5.7 恒等函数 如果 $I_x=\{<x,x>|x\in X\}$，则称函数 $I_x: X\to X$ 为恒等函数。

5.3 函数的运算(复合、逆函数)

函数是一种特殊的关系，自然也能进行关系所特有的运算，如关系的复合与关系的逆等。但由于函数定义的特殊要求，在进行类似复合及逆运算时，会有相关的条件限制。

5.3.1 复合函数

定理 5.2 对于集合 X、Y 和 Z，设 f 是 X 到 Y 的关系、g 是 Y 到 Z 的关系，如果 f 和 g 分别是 X 到 Y 的函数和 Y 到 Z 的函数，那么，其复合关系 $f\circ g$ 是 X 到 Z 的函数。

【证明】 本定理可从函数定义的两个要素出发进行证明。

① 因为 f 为函数，所以对于每一个 $x\in X$，必存在一个 $y\in Y$，使得 $<x,y>\in f$；同理，因为 g 为函数，所以对于每一个 $y\in Y$(包括 x 在 Y 的像点)，必存在一个 $z\in Z$，使得 $<y,z>\in g$。因此，根据复合关系的概念可知，对于每一个 $x\in X$，必存在一个 $z\in Z$，使得 $<x,z>\in f\circ g$，即经过关系的复合后，集合 X 中的每一个元素，在集合 Z 中都有像点；

② 假定对于某个 $x\in X$，经关系的复合运算后，在集合 Z 中存在两个不同的像点 z_1、$z_2(z_1\neq z_2)$，即 $<x,z_1>\in f\circ g$、$<x,z_2>\in f\circ g$。根据关系复合运算的规则可知，必有 $<x,y_1>\in f$、$<y_1,z_1>\in g$、$<x,y_2>\in f$、$<y_2,z_2>\in g$。由于 f 为函数，则 X 中的每个元素仅有唯一的像，因此，必有 $y_1=y_2$。同样，由于 g 为函数，则有 $z_1=z_2$，这与假设 $z_1\neq z_2$ 相矛盾，所以假设不成立，即对于任意 $x\in X$，经过关系的复合运算后，在集合 Z 中只能有唯一的像点。

综合以上两点，本定理得证。

定义 5.8 复合函数 对于集合 X、Y 和 Z，设函数 $f:X\to Y$，$g:Y\to Z$，则复合关系 $f\circ g$ 称为函数 f 与函数 g 的复合函数。为了与其他资料上复合函数的表示方式一致，通常将复合函数记为 $g\circ f$。因此，函数的复合也称为函数 g 在 f 的左边可复合。

函数复合运算的数学描述形式为：

$$g\circ f=\{<x,z>|x\in X\wedge z\in Z\wedge(\exists y)(y\in Y\wedge y=f(x)\wedge z=g(y))\}$$

根据复合函数的定义，显然有 $g\circ f(x)=g(f(x))$。

例题 5.5 设集合 $X=\{1,2,3\}$、$Y=\{p,q\}$、$Z=\{a,b\}$，函数 $f=\{<1,p>,<2,p>,<3,q>\}$、$g=\{<p,b>,<q,b>\}$，求复合函数 $g\circ f$。

【解】 $g\circ f=\{<1,b>,<2,b>,<3,b>\}$

例题 5.6 设 R 为实数的集合，对 $x\in R$ 有 $f(x)=x+2$、$g(x)=x-2$、$h(x)=3x$，试求 $g\circ f$、$h\circ g$、$h\circ(g\circ f)$、$(h\circ g)\circ f$。

【解】 因为 $g\circ f(x)=g(f(x))=g(x+2)=x$，所以 $g\circ f=\{<x,x>|x\in R\}$；

同理可得

$$h\circ g=\{<x,3x-6>|x\in R\}$$

$$h\circ(g\circ f)=\{<x,3x>|x\in R\}$$

$$(h\circ g)\circ f=\{<x,3x>|x\in R\}$$

根据例题5.6的解答可以看出，函数的复合运算是可以结合的，即 $h\circ(g\circ f)=(h\circ g)\circ f$，因此，在多个函数的复合时，可以去掉上式中的括号。

定理5.3 设 $f:X\to Y$，$g:Y\to Z$，$g\circ f$ 是一个复合函数，则：

① 如果 f 和 g 是满射的，则 $g\circ f$ 也是满射的。

② 如果 f 和 g 是入射的，则 $g\circ f$ 也是入射的。

③ 如果 f 和 g 是双射的，则 $g\circ f$ 也是双射的。

【证明】

① 设有任意的元素 $z\in Z$。因为 g 是满射的，所以 z 必为 Y 中某个元素 y 的像，即必有某个 $y\in Y$ 使得 $g(y)=z$；同样的，因为 f 是满射的，所以每一个 $y\in Y$ 必为 X 中某个元素 x 的像，即必有某个元素 $x\in X$ 使得 $f(x)=y$。所以，对于任意的元素 $z\in Z$，其必为 X 中某个元素的像，即 $g\circ f(x)=g(f(x))=g(y)=z$。

因此，$g\circ f$ 是满射的。

② 令 x_1、x_2 为 X 的元素，且 $x_1\neq x_2$。因为 f 是入射的，所以 $f(x_1)\neq f(x_2)$。同样地，因为 g 是入射的，则 $g(f(x_1))\neq g(f(x_2))$。于是有 $x_1\neq x_2\Rightarrow g\circ f(x_1)\neq g\circ f(x_2)$，所以，$g\circ f$ 也是入射的。

③ 因为 g 和 f 是双射，故根据①和②可知，$g\circ f$ 也是满满射和入射的，即 $g\circ f$ 是双射的。

本定理证毕。

由于2个函数复合后仍然是函数，因此，函数的复合运算可以推广到3个甚至多个函数。

5.3.2 逆函数

定理5.4 如果关系 $f:X\to Y$ 是一个双射函数，那么其逆关系 f^c 是 Y 到 X 的双射函数。

【证明】 设 $f=\{<x,y>|x\in X\wedge y\in Y\wedge f(x)=y\}$、$f^c=\{<y,x>|<x,y>\in f\}$。

① 因为 f 是双射函数，所以 f 既是满射的又是入射的。如果 f 是满射的，则对于每一个 $y\in Y$，都有 x 与它对应；如果 f 是入射的，则对于任意的 $y\in Y$，都有唯一的 x 与它对应。所以 f^c 是一个函数。

② 由于 f 是函数，则对于每一个 x 都有一个 y 与之对应，所以 f^c 是一个满射函数。

③ 若 f^c 不是入射函数，则必存在 $y_1\neq y_2$，使得 $f^c(y_1)=f^c(y_2)$。令 $f^c(y_1)=x_1$、$f^c(y_2)=x_2$，可得 $x_1=x_2$，于是有 $f(x_1)=f(x_2)$。由 f^c 是 f 的逆关系可知 $y_1=f(x_1)$、$y_2=f(x_2)$，于是可得 $y_1=y_2$。这与前提条件 $y_1\neq y_2$ 相矛盾，所以 f^c 只能是入射函数。

综合以上3点，本定理得证。

定义5.9 逆函数 设 $f:X\to Y$ 是一个双射函数，称 $Y\to X$ 的双射函数 f^c 为 f 的逆函数，记为 f^{-1}。

例题5.7 设 $X=\{1,2,3\}$、$Y=\{a,b,c\}$，$f:X\to Y$ 为 $f=\{<1,a>,<2,c>,<3,b>\}$，求 f^{-1}。

【解】 由题目可知 f 为双射函数，所以有逆函数 f^{-1}。

$$f^{-1}=\{<a,1>,<c,2>,<b,3>\}$$

如果例题 5.3 的函数为 $f=\{<1,a>,<2,b>,<3,b>\}$，因为其不是双射函数，则其逆关系 $f^c=\{<a,1>,<b,2>,<b,3>\}$不是函数。

定理 5.5 设 $f:X\to Y$，则 $f=f\circ I_x=I_y\circ f$

本定理可以根据复合函数和恒等关系(恒等函数)的定义进行证明。

定理 5.6 如果函数 $f:X\to Y$ 有逆函数 $f^{-1}: Y\to X$，则有 $f^{-1}\circ f=I_x$ 且 $f\circ f^{-1}=I_y$。

【证明】

① 因为 f 是双射函数，所以 f^{-1} 也是双射函数。

② $f^{-1}\circ f$ 与 I_x 的定义域都是 X；$f\circ f^{-1}$ 与 I_y 的值域都是 Y。

令 $y=f(x)$、$x=f^{-1}(y)$，则 $f^{-1}\circ f(x)=f^{-1}(y)=x$、$f\circ f^{-1}(y)=f\circ f^{-1}(x)=y$，即 $f^{-1}\circ f=I_x$、$f\circ f^{-1}=I_y$。

本定理证毕。

例题 5.8 令 f：$\{1,2,3\}\to\{a,b,c\}$的定义如图 5.2(a)所示，试求 $f^{-1}\circ f$ 与 $f\circ f^{-1}$。

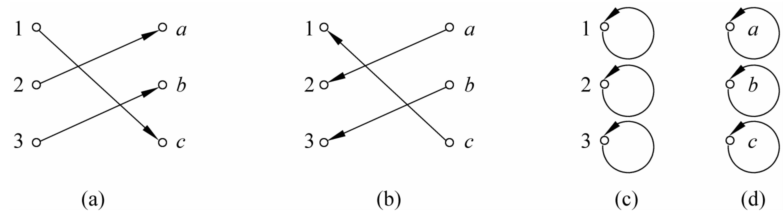

图 5.2 例题 5.8 图

【解】 f^{-1}可表示为图 5.2(b)；

$f^{-1}\circ f$ 可表示为图 5.2(c)；

$f\circ f^{-1}$可表示为图 5.2(d)。

定理 5.7 若 $f:X\to Y$ 是一一对应的函数，则$(f^{-1})^{-1}=f$。

【证明】

① 因为 $f:X\to Y$ 是一一对应的函数，所以 $f^{-1}: Y\to X$ 也是一一对应的函数。

同理，$(f^{-1})^{-1}: X\to Y$ 又是一一对应的函数。于是有：

$$\mathrm{dom}\ f=\mathrm{dom}\ (f^{-1})^{-1}=X$$

② 设有任意的 $x\in X$，则：

$$\begin{aligned} x\in X &\Rightarrow f: x\to f(x) \\ &\Rightarrow f^{-1}: f(x)\to x \\ &\Rightarrow (f^{-1})^{-1}: x\to f(x) \end{aligned}$$

由①和②可得$(f^{-1})^{-1}=f$，本定理证毕。

定理 5.8 设 $f:X\to Y$、$g:Y\to Z$ 都是一一对应(双射)的函数，则$(g\circ f)^{-1}=f^{-1}\circ g^{-1}$。

【证明】

① 根据定理 5.4 可知，因为 $f:X\to Y$、$g:Y\to Z$ 都是双射函数，所以 f^{-1}、g^{-1}存在且 $f^{-1}: Y\to X$、$g^{-1}: Z\to Y$ 是双射函数，再根据定理 5.3 可知 $f^{-1}\circ g^{-1}: Z\to X$ 是双射函数。

同样，根据定理 5.4 可知，如果 $f:X\to Y$、$g:Y\to Z$ 都是双射函数，则 $g\circ f:X\to Z$ 也是

双射函数，故$(g\circ f)^{-1}$存在且$(g\circ f)^{-1}: Z\to X$是双射函数。

综上可知，$f^{-1}\circ g^{-1}$与$(g\circ f)^{-1}$都是$Z\to X$的双射函数。

② 因$f^{-1}: Y\to X$、$g^{-1}: Z\to Y$是双射函数，则对于任意$z\in Z$，存在唯一的$y\in Y$，使得$g^{-1}(z)=y$，且存在唯一的$x\in X$，使得$f^{-1}(y)=x$。于是有：

$$(f^{-1}\circ g^{-1})(z)=f^{-1}(g^{-1}(z))=f^{-1}(y)=x$$

又因为

$$(g\circ f)(x)=g(f(x))=g(y)=z$$

则

$$(g\circ f)^{-1}(z)=x$$

因此对任意$z\in Z$有：

$$(g\circ f)^{-1}(z)=(f^{-1}\circ g^{-1})(z)$$

综合①②可知，如果$f: X\to Y$、$g: Y\to Z$都是双射函数，则$(g\circ f)^{-1}=f^{-1}\circ g^{-1}$。

本定理证毕。

5.4　函数的应用

哈希函数是广义上函数的一种实际应用。

哈希函数也称为散列函数或 Hash 函数，是一种将任意长度的输入字符串转化成固定长度字符串输出的函数。在数据结构中，常利用哈希函数来加速对数据项的查找过程。根据应用情况的不同，有时也将哈希函数的输出值称为哈希值、哈希码、散列值等。

在线性表、树等数据结构中，每个记录所在的相对位置是随机的，与记录的关键字之间不存在确定关系。因此，在这些数据结构中查找记录时需要进行一系列的关键字比较，而查找效率的高低则依赖于查找过程中所进行的比较次数。如果希望不经过任何比较，仅用一次存取就能得到需要的记录，那么就必须建立一个关于存储位置和关键字的对应关系函数f，使得每个关键字与一个唯一的存储位置相对应。在查找时，只要根据这个关系f就能找到给定值key的像$f(key)$。如果数据结构中存在关键字与key相等的记录，则其必定在$f(key)$的存储位置上，这样不需要进行比较就可以直接取得所要查找的记录。这里的对应关系f实际上就是一个哈希函数，而按照这种思路建立起来的表格则称为哈希表。

例如，如果要建立一张学生成绩表，最简单的方法是以学生的学号作为关键字，1号学生的记录位置在第1条，2号学生的记录位置在第2条，依此类推。此时，如果要查看学号为5的学生的成绩，则只需要取出第5条记录就可以。这样建立的表实际上就是一张简单的哈希表，其哈希函数为$f(key)=key$。

然而，很多情况下的哈希函数并不如此简单。为了查看方便，有时可能会以学生的名字作为关键字。此时，为了能够根据学生的名字直接定位出相应记录所在的位置，需要将这些名字转化为数字，再构造出相应的哈希函数。以下是两不同的哈希函数。

① 考查学生名字的汉语拼音，将其中第一个字母在英语字母表中的序号作为哈希函

数的值。例如，某学生“蔡军”，其汉语拼音的第一个字母为C，因此可取03作为其哈希值。

② 考查学生名字的汉语拼音，将其中第一个字母和最后一个字母在英语字母表中的序号之和作为哈希函数的值。同样的，对于“蔡军”这个名字，其汉语拼音的第一个字母和最后一个字母分别为C和N，因此取3+14=17作为其哈希值。

分别应用这两个哈希函数，成绩表中部分学生名字不同的哈希函数值如表5-1所示。

表5-1 部分学生名字的哈希函数值

key	李丽 (LILI)	赵宏英 (ZHAOHONGYING)	肖军 (XIAOJUN)	吴小艳 (WUXIAOYAN)	肖秋梅 (XIAOQIUMEI)	陈伟 (CHENWEI)
$f_1(key)$	12	26	24	23	24	03
$f_2(key)$	21	33	38	37	33	12

从表5-1可以看出，在哈希表的构造过程中，可能会出现不同的关键字映射到同一地址的情况，即 $key_1 \neq key_2$ 但 $f(key_1)=f(key_2)$，这种现象称为冲突或碰撞。实际上，由于哈希函数是把任意长度的字符串映射为固定长度的字符串，冲突是必然存在的。因此，冲突不可能避免，只能尽可能减少。例如上面给出的两个哈希函数中，应用第二个函数时出现的碰撞就比第一个函数要少得多。

为了减少建立函数时出现碰撞的概率，实际应用中，常见构造哈希函数的方法一般有以下几种：

1. 直接定址法

直接定址法取关键字或关键字的某个线性函数值为哈希地址，即 $f(key)=key$ 或 $f(key)=a \cdot key+b$，其中 a 和 b 为常数。直接定址法所得到的地址空间与关键字集合的大小相同，对于不同的关键字不会发生冲突，但在实际应用中使用这种哈希函数的情况比较少。

2. 数字分析法

数字分析法适合于关键字由若干数码组成，且各数码的分布规律预先知道的情况。具体方法是：分析关键字集合中每个关键字的每一位数码的分布情况，找出数码分布均匀的若干位作为关键字的存储地址。

例如，一个由80个结点组成的结构，其关键字为6位十进制数。选择哈希表长度为100，则可取关键字中的两位十进制数作为结点的存储地址。具体采用哪两位数码，需要用数字分析法对关键字中的数码分布情况进行分析。假设结点中有一部分关键字为：

$$key_1=301514 \qquad key_2=303027$$
$$key_3=301103 \qquad key_4=308329$$
$$key_5=300287 \qquad key_6=305939$$
$$key_7=300792 \qquad key_8=300463$$

对上述关键字分析可以发现，关键字的第1位均为3、第2位均为0，分布集中，不适

合作为存储地址。而第4位和第5位分布均匀，所以该哈希函数可以构造为取第4、5位作为结点的存储地址。上述8个结点的哈希地址分别为：

$f(key_1)=51$　　$f(key_2)=02$　　$f(key_3)=10$　　$f(key_4)=32$

$f(key_5)=28$　　$f(key_6)=93$　　$f(key_7)=79$　　$f(key_8)=46$

3. 平方取中法

平方取中法是一种比较常用的构造哈希函数的方法，其具体方法是：将关键字求平方，然后取中间的几位数字作为哈希地址。由于关键字求平方后的中间几位数字和组成关键字的每一位数字都有关，因此产生冲突的可能性相对较小，最后究竟取几位数字作为散列地址需要由散列表的长度决定。

例如，若数据结构的存储地址范围是1～999，则取平方值的中间三位，如表5-2所示。

表5-2　平方取中法

关键字 key	key^2	哈希地址	压缩地址
11032710	121720689944100	689	344
11054312	122197813793344	813	406
01110345	001232866019025	866	433
01111401	001235212182801	212	106

若所取哈希函数值超出了存储区的地址范围，则可以再乘以一个比例因子，把哈希函数值放大或缩小，使其位于存储区的范围内。如果上述示例中存储地址范围仅是1～500，则原哈希值会超过地址范围，此时可以对哈希地址再乘以0.5取整(表5-2的压缩地址)。

4. 叠加法

折叠法适用于关键字位数很多且关键字中每一位上数字分布大致均匀的情况。具体方法是：将关键字分割(或折叠)成位数相同的几部分(最后一部分的位数可以不同)，然后将这几部分的相加的和(舍去向最高位的进位)作为哈希地址。叠加的方法有两种：移位叠加和间界叠加。移位叠加是将分组后的每组数字的最低位对齐，然后相加；间界叠加是将分组后的每组数字从一端向另一端沿分界线进行来回折叠，然后对齐相加。

例如，西文图书的国际标准图书编号是一个10位的十进制数，对某图书编号为0-383-40284-6，采用4位叠加。对于移位叠加，则从最低位开始取4位"2846"，左移4位相加(即2846+8340=11186)，相加结果再左移4位相加(即11186+03=11189)，最后取相加结果的低4位(即去掉进位位)"1189"为哈希地址，$f(key)=1189$。

间界叠加则是：从最低位开始取4位进行折叠，由于图书编号为10位，因此需折叠两次，折叠完后，对应位相加(即2846+0438+03=3287)，最后取相加结果的低4位(无进位位，不需去掉)"3287"为哈希地址，$f(key)=3287$。

5. 除留余数法

除留余数法取关键字被某个不大于哈希表表长 m 的数 p 除后所得余数为哈希地址，

即 $f(key)=key(\bmod p)$，其中 $p\leqslant m$，这是构造哈希函数的最简单也是最常用的一种方法。它不仅可以对关键字直接取模(mod)，而且可在折叠、平方取中等运算之后再取模。值得注意的是，在使用除留余数法时，对 p 的选择很重要。若 P 选得不好，容易产生相同的哈希值。根据经验可知：一般情况下可以选 p 为质数，有时也可以选取没有小于 20 的质素的合数。

6. 随机数法

随机数法是选择一个随机函数，再取关键字的随机函数值作为它的哈希地址，即 $f(key)=\mathrm{random}(key)$。其中，$\mathrm{random}(x)$ 为随机函数。随机数法适用于关键字长度不等时，构造哈希函数。

由于构造哈希函数时，很难避免的出现冲突的情况，因此在构建哈希表时不仅要设定一个"好"的哈希函数，而且还要设定一种处理冲突的方法。假设哈希表的地址集为 0～$(n-1)$，冲突是指由关键字得到的哈希地址为 $j(0\leqslant j\leqslant n-1)$ 的位置上已存有记录，则"处理冲突"就是为该关键字的记录找到另一个"空"的哈希地址。在处理冲突的过程中可能得到一个地址序列 h_i，其中，$h_i\in[0,n-1]$，$i=1,2,\cdots,k$。即在处理哈希地址的冲突时，若得到的另一个哈希地址 h_1 仍然发生冲突，则再求下一个地址 h_2，若 h_2 仍然冲突，再求得 h_3，依此类推，直至 h_k 不发生冲突为止，则 h_k 为该记录在表中的地址。

通常处理冲突的方法有下列几种：

1. 开放定址法

开放定址法是当冲突发生时，形成一个检测序列，然后沿此序列逐个进行地址检测，直到找到一个空位置(开放的地址)，将发生冲突的记录放到该地址中，即 $h_i=(f(key)+d(i)(\bmod m)$，$i=1,2,\cdots,k(k\leqslant m-1)$，其中，$f(key)$为哈希函数，$m$ 为哈希表表长，$d(i)$为增量序列。根据对 $d(i)$的设置情况，开放定址法又可以有以下三种不同的方法：

① 线性检测再散列。$d_i=1,2,3,\cdots,m-1$；

② 二次检测再散列。$d_i=1^2,-1^2,2^2,-2^2,3^2,\cdots,\pm k^2,(k\leqslant m/2)$；

③ 伪随机检测再散列。d_i=伪随机数序列。

2. 再哈希法

再哈希法是当不同关键字的映射地址产生冲突时，通过另一个哈希函数计算地址，直到冲突不再发生，这种方法不易产生"聚集"，但增加了计算的时间。

3. 拉链法

拉链法是将所有关键字映射地址有冲突的记录存储在同一个线性链表中。假设某哈希函数产生的哈希地址在区间$[0,m-1]$上，则首先设立一个指针型向量 ChainHash[m]，其每个分量的初始状态都是空指针。接下来，对于哈希地址为 i 的所有记录都插入到以 ChainHash[i]为头指针的链表中。

例如，已知一组关键字为(26、36、41、38、44、15、68、12、06、51)，表长为 13，选定的哈希函数为 $f(key)=key(\bmod 13)$。显然关键字 41、15 的哈希地址都为 2，关键字 38、12、51 的哈希地址都为 12，产生了映射地址的冲突。此时，可选用拉链法解决这一问题，由于