Оценки для хеширования

13 декабря 2021 г. 13:13

Def. Семейство хеш-функций – мн-во ф-ий $\mathcal{H}=\{h_i\}_{i\in I}$ $\quad \forall h_i : A \to B$

Использование: в начале прогр. один раз выбирают $h \leftarrow U(\mathcal{H})$

Пример: полин. хеширование $\mathcal{H}=\{h_{m,x}\}_{x\in\mathbb{Z}_m}$ (m – фикс. простое число) $\quad h_{m,x}: \Sigma^* \to \mathbb{Z}_m$

Def. Семейство х.-ф. $\mathcal{H}$ универсально $\overset{\text{def}}{\Leftrightarrow}$ $\forall a_1,a_2\in A,\ a_1\neq a_2 \Rightarrow \Pr_{h\leftarrow U(\mathcal{H})}[h(a_1)=h(a_2)] \le \frac{1}{|B|}$ (вероятн. коллизии)

Th. Хеш-табл. с закрытой адресацией, исп. унив. семейство х.-ф. $\mathcal{H}$.
$n \le m\cdot c$, m – размер хеш-таблицы, n – кол-во эл-ов
Тогда математическое ожидание времени работы каждой операции есть $O(1+c)$.

Д-во Пусть в х-т. уже есть $x_1,\dots,x_n$ (различные). Совершаем операцию с x_{n+1}.
Докажем, что $E[\text{размера списка, в кот. попадёт } x_{n+1}] \le O(1+c)$

$$E_{h\leftarrow U(\mathcal{H})}[\text{рассм. сп.}] = E\Big[\sum_{i=1}^{n} \mathbb{1}_{h(x_i)=h(x_{n+1})}\Big] = \sum_{i=1}^{n} E\big[\mathbb{1}_{h(x_i)=h(x_{n+1})}\big] =$$

$$= \sum_{i=1}^{n} \Pr[h(x_i)=h(x_{n+1})] = \sum_{\substack{i=1\\ x_i\neq x_{n+1}}}^{n} \underbrace{\Pr[h(x_i)=h(x_{n+1})]}_{\le |B|} + \sum_{\substack{i=1\\ x_i= x_{n+1}}}^{n} \underbrace{\Pr[h(x_i)=h(x_{n+1})]}_{=1} \le$$

$$\le \frac{(n-1)}{|B|} + 1 = \frac{n-1}{m} + 1 \le c+1$$

$\mathbb{1}_P = \begin{cases} 1, & P \text{ вып.} \\ 0, & P \text{ не вып.} \end{cases}$ $\quad E[\mathbb{1}_P] = \Pr[P \text{ вып.}]$

$|B| = m$, $k = h(x_i)$ – индекс от 0 до $m-1$

■

Полином. хеширование

$$h_{m,x}(s) = \left(\sum_{i=0}^{|s|-1} x^{|s|-i-1}\cdot s_i\right) \bmod m$$

① Строки случайные

Утв. $x_0\in\mathbb{Z}_m$
P – случайный многочлен $\deg<n$ с коэфф. из $\mathbb{Z}_m$ $\quad P(x)=a_0+a_1x+a_2x^2+\dots+a_{n-1}x^{n-1}$
$a_0,\dots,a_{n-1}$ – случ. независ. р/вер. из $\mathbb{Z}_m$

Тогда $\Pr_P[P(x_0)\equiv_m 0] = \frac{1}{m}$.

Д-во Сначала сгенер. $a_{n-1},\dots,a_1$. $P(x_0) = \underbrace{a_1x_0+\dots+a_{n-1}x_0^{n-1}}_{t} + a_0$

Генерируем a_0. $\Pr_{a_0\leftarrow U(\mathbb{Z}_m)}[t+a_0\equiv_m 0] = \frac{1}{m}$. ■

Lm. $\forall m\in\mathbb{P}$, $x_0\in\mathbb{Z}_m$. Тогда $\Pr_{s,t\text{ – случ. строки длины } n}[h_{m,x_0}(s)=h_{m,x_0}(t)] = \frac{1}{m}$

Д-во. $h_{m,x_0}(s) = S(x_0)$, S – многочлен (mod m), $\deg S<n$, S выбир. случ.
$h_{m,x_0}(t) = T(x_0)$, T – многочл. (mod m), – " –

$$\Pr[h_{m,x_0}(s)=h_{m,x_0}(t)] = \Pr_{\substack{S,T\text{ – сл. мн-ны.}\\ \deg S,\deg T<n}}[S(x_0)\equiv_m T(x_0)] = \Pr_{S,T}[\underbrace{(S-T)}_{\text{случ. мн.-н.}}(x_0)\equiv_m 0] = \Pr_{P\text{ – сл. мн.}}[P(x_0)\equiv_m 0] = \frac{1}{m}$$

■

k сравнений $\Rightarrow \Pr[\text{произошла} \ge 1 \text{ колл.}] \le k\cdot\Pr[\text{коллизия}] \le \frac{k}{m}$

$\Pr[\text{колл.}_1 \vee \text{колл.}_2 \vee \text{колл.}_3 \vee \dots \vee \text{колл.}_k]$

② x_0 выбираем случайно из $\mathbb{Z}_m$ $\quad m\in\mathbb{P}$

Утв. P – многочлен над $\mathbb{Z}_m$, $\deg P\le n$. Тогда P имеет $\le n$ корней (mod m). $(P\neq 0)$
(следств. из теоремы Безу)

Тут важно, что $m\in\mathbb{P}$. Пример: пусть $m=2^{64}$. $P(x)=x^{64}$ $\deg P=64$. Имеет 2^{63} корней – любое чётное число $0,\dots,m-1$

Lm. s,t – строки, $s\neq t$, $|s|,|t|\le n$ $\quad \Pr_{x_0\leftarrow U(\mathbb{Z}_m)}[h_{m,x_0}(s)\equiv_m h_{m,x_0}(t)] \le \frac{n}{m}$

Д-во.

$$= \Pr_{S,T\text{ – мн-ны.}}[S(x_0)\equiv_m T(x_0)] = \Pr[\underbrace{(S-T)}_{P}(x_0)\equiv_m 0] = \Pr[\underbrace{P(x_0)}_{\text{фикс.}}\equiv_m 0] =$$

$$= \frac{\#\{x_0 : P(x_0)\equiv_m 0\}}{\#\{x_0\}} \le \frac{n}{m} \quad (P\neq 0, \text{т.к. } s\neq t)$$

■

k сравнений $\Rightarrow \Pr[\exists \ge 1 \text{ колл.}] \le \frac{kn}{m}$