

連想値を用いるハッシング関数について

玉越, 靖司
九州大学大学院総合理工学研究科情報システム学専攻

有川, 節夫
九州大学大学院総合理工学研究科情報システム学専攻

<https://doi.org/10.15017/17556>

出版情報 : 九州大学大学院総合理工学報告. 4 (1), pp.79-86, 1982-06-30. 九州大学大学院総合理工学
研究科
バージョン :
権利関係 :

連想値を用いるハッシング関数について

玉越靖司*・有川節夫**

昭和57年3月29日 受理

On Hashing Functions Using Associated Values Yasushi TAMAKOSHI and Setsuo ARIKAWA

A new type of hashing functions which make use of associated values of keys is introduced. Necessary and sufficient conditions of minimality and perfectness of such functions are presented. A similar condition for Cichelli's hashing functions is also presented. The time complexities of deciding the perfectness and the minimality are discussed. The cost of our functions is numerically evaluated.

1. はじめに

ハッシュ法は、キー比較コストが定数に近い探索法である。特に、計算機システムのコマンドのような特定のキーの集合に対して、その集合の1つの要素について探索する場合、非常に効果的である。このような静的なキーの集合に対しては、ハッシング関数を工夫することによって、キー比較コストを減少させることが可能である。そこで、望ましいハッシング関数として、キー比較回数が1つのキーに対してただ1回だけである完全ハッシング関数 (perfect hashing function)¹⁾ が考えられる。

完全ハッシング関数の1つの方法として、連想値 (associated value) を用いる方法²⁾ がある。この方法は、関数値の計算に割り算を含まずに整数の足し算だけでキーのハッシュ値を計算するもので、計算機での実行に際して演算時間が短くなるという利点を持つ。しかしながら、この方法では Jaeschke ら³⁾ の指摘を待つまでもなく、完全ハッシング関数を持たないようなキーの集合が存在する。実用的見地からは、キーの集合が与えられたとき、それに対して完全ハッシング関数が存在するか否かを判定することは極めて重要である。そこで本稿では、まず、与えられたキーの集合に対して、連想値を用いる完全ハッシング関数が存在するための必要十分条件について考察する。

完全ハッシング関数が存在すれば、ハッシュテーブルの大きさ M はキーの集合の要素の数 n 以上であるが、このとき、ハッシュテーブルのうちで $M-n$ 個のバケットは使われていない。格納率 (loading factor; n/M) が1のとき、その関数は最小完全ハッシング関数 (minimal perfect hashing function)¹⁾ と呼ばれ、計算機の記憶域コストの面からは理想的である。こうした連想値を用いる最小完全ハッシング関数に対する存在性の判定問題について考察し、そのアルゴリズムを与える。

§2では、我々が定義した型⁴⁾ の連想値を用いるハッシング関数について、その完全な関数の存在条件および格納率に関する問題、最小完全な関数の存在性に関する決定問題等について述べる。§3では、Cichelli²⁾ が定義した型の関数について、§2と同様の議論が可能であることを述べ、完全な関数の存在条件を示す。また、Cichelli のハッシング関数は、我々が定義した型のハッシング関数の族に含まれることを示す。さらに§4では、関数のコストを計算機による関数値計算の演算時間の面から、連想値を用いないハッシング関数と比較して検討する。

2. 連想値を用いる完全ハッシング関数

本節では、まず、連想値を用いるハッシング関数の定義を与え、完全な関数が存在するための必要十分条件を示す。次に、連想値の割り当て方と格納率の問題について述べ、最後に、与えられたキーの集合に対して最小完全な関数の存在性を決定することが可能であ

* 情報システム学専攻修士課程

** 情報システム学専攻

ることを示す。

2.1. 諸定義

以下の議論に必要な文字の有限集合を, $\Sigma = \{r_1, r_2, \dots, r_m\}$ とし, $I = \{w_1, w_2, \dots, w_n\} \subset \Sigma^l$ をキーの集合と呼ぶ。I の要素をキーと呼ぶ。

〔注意1〕

I に属するすべてのキーの長さが等しいと仮定しても一般性を失わない。それは、長さの異なるキーの集合に対しても、その最長のキーの長さ I より短いキーには_ (空白) などの文字を埋めて、長さを I にすることができるからである。

〔定義1〕

文字 $x, y \in \Sigma$ に対して

$$d(x, y) \equiv \begin{cases} 0 & (x = y) \\ 1 & (x \neq y). \end{cases}$$

〔定義2〕

キー $u, v \in I$ に対して

$$u = v \iff \sum_{k=1}^l d(u^{(k)}, v^{(k)}) = 0$$

ここに、 $w^{(i)}$ は、キー w の i 番目の文字とする。

〔定義3〕

キー $w \in I$ のうち p 個の文字; $w^{(i_1)}, w^{(i_2)}, \dots, w^{(i_p)}$ ($1 \leq p \leq l, s < t \Rightarrow 1 \leq i_s < i_t \leq l$) の中で、 $r \in \Sigma$ と等しい文字の個数を次のように書く。

$$|w|_{r^{(i_1), \dots, i_p}} = p - \sum_{k=1}^p d(w^{(i_k)}, r).$$

〔例1〕

$w = \text{ABCBC}$ とする ($l = 5$).

$$|w|_{A^{(1,3,5)}} = 1, |w|_{B^{(1,3,5)}} = 0, |w|_{C^{(1,3,5)}} = 2.$$

〔定義4〕

$h: I \rightarrow N$ の1対1写像: I に対する完全ハッシング関数と呼ぶ。ここに、N は0以上の整数の全体である。

〔定義5〕

I に対する完全ハッシング関数 h が

$$\max h(I) - \min h(I) = n - 1$$

をみたすとき、 h を I に対する最小完全ハッシング関数と呼ぶ。

〔定義6〕

$a: \Sigma \rightarrow Z: a$ を連想といい、 $a(r)$ を文字 r の連想値と呼ぶ。ここに、Z は整数の全体である。

〔定義7〕

a を連想とすると、

$$h(w) = a(w^{(i_1)}) + a(w^{(i_2)}) + \dots + a(w^{(i_p)}) \quad (w \in I)$$

で定義される関数 $h: I \rightarrow N$ を、連想値を用いるハッシング関数といい、今後、簡単に Aハッシング関数と呼ぶことにする。

〔例2〕

キーの集合として、英語による暦月名を考える。

$$\text{Function: } h(w) = a(w^{(2)}) + a(w^{(3)})$$

Associated Values:

$$\begin{aligned} a(A) &= 4 & a(B) &= 5 & a(C) &= 2 & a(D) &= 0 \\ a(E) &= 0 & a(F) &= 0 & a(G) &= 3 & a(H) &= 0 \\ a(I) &= 0 & a(J) &= 0 & a(K) &= 0 & a(L) &= 6 \\ a(M) &= 0 & a(N) &= 0 & a(O) &= 5 & a(P) &= 1 \\ a(Q) &= 0 & a(R) &= 6 & a(S) &= 0 & a(T) &= 6 \\ a(U) &= 0 & a(V) &= 6 & a(W) &= 0 & a(X) &= 0 \\ a(Y) &= 5 & a(Z) &= 0 & a(_) &= 0 \end{aligned}$$

I (Set of keys)

Hash Values

$w_1 = \text{JANUARY}$ _	$h(w_1) = 4$
$w_2 = \text{FEBRUARY}$ _	$h(w_2) = 5$
$w_3 = \text{MARCH}$ _____	$h(w_3) = 10$
$w_4 = \text{APRIL}$ _____	$h(w_4) = 7$
$w_5 = \text{MAY}$ _____	$h(w_5) = 9$
$w_6 = \text{JUNE}$ _____	$h(w_6) = 0$
$w_7 = \text{JULY}$ _____	$h(w_7) = 6$
$w_8 = \text{AUGUST}$ _____	$h(w_8) = 3$
$w_9 = \text{SEPTEMBER}$ _____	$h(w_9) = 1$
$w_{10} = \text{OCTOBER}$ _____	$h(w_{10}) = 8$
$w_{11} = \text{NOVEMBER}$ _____	$h(w_{11}) = 11$
$w_{12} = \text{DECEMBER}$ _____	$h(w_{12}) = 2.$

この場合は、最小完全な Aハッシング関数となっている。

2.2. 完全な関数の存在条件

Aハッシング関数では、完全ハッシング関数を存在させ得ないようなキーの集合 I がある (例3)。

〔例3〕

キーの集合として、 $I = \{AA, AB, BA\}$ ($l = 2, w_1 = AA, w_2 = AB, w_3 = BA$) を考える。

(証明)

i) $p = 1, i_1 = 1$ のとき、

$$h(w_1) = h(w_2) = a(A) \text{ となる。}$$

ii) $p=1, i_1=2$ のとき,

$$h(w_1) = h(w_3) = a(A) \text{ となる.}$$

iii) $p=2, i_1=1, i_2=2$ のとき,

$$h(w_2) = h(w_3) = a(A) + a(B) \text{ となる.}$$

従って, 完全な A ハッシング関数は存在しない□

次に, 与えられたキーの集合に対して, A 完全ハッシング関数が存在するための必要十分条件を示す.

〔定理 1〕

キーの集合 I に対して, 完全な A ハッシング関数が存在する.

$$\Leftrightarrow \exists p, i_1, \dots, i_p : i) 1 \leq p \leq l$$

$$ii) s < t \Rightarrow 1 \leq i_s < i_t \leq l$$

$$iii) \text{ 任意の } u, v \in I, u \neq v \text{ に対し}$$

て

$$\sum_{k=1}^m ||u|_{r_k}^{(i_1, \dots, i_p)} - |v|_{r_k}^{(i_1, \dots, i_p)}| \neq 0.$$

〔証明〕

◁ $a(r_j) = (p+1)^{j-1} (1 \leq j \leq m)$ とすると,

$h(w) = |w|_{r_1}^{(i_1, \dots, i_p)} \times 1 + |w|_{r_2}^{(i_1, \dots, i_p)} \times (p+1) + \dots + |w|_{r_m}^{(i_1, \dots, i_p)} \times (p+1)^{m-1}$ となり, $0 \leq |w|_{r_j}^{(i_1, \dots, i_p)} \leq p$ より, $h(w)$ は, m 桁の $(p+1)$ 進数として表わされる. しかも,

$$\sum_{k=1}^m ||u|_{r_k}^{(i_1, \dots, i_p)} - |v|_{r_k}^{(i_1, \dots, i_p)}| \neq 0$$

より, 任意の, I の異なる 2 要素 u, v に対して

$$h(u) \neq h(v)$$

となる.

⇒ 完全な A ハッシング関数が存在する.

◁ ⇒ $\exists a(r_1), \dots, a(r_m) : \forall u, v \in I, u \neq v$ に対して

$$i) h(w) = a(w^{(i_1)}) + a(w^{(i_2)}) + \dots + a(w^{(i_p)})$$

$$(w \in I)$$

$$ii) h(u) \neq h(v)$$

より明らか□

〔注意 2〕

ある p, i_1, \dots, i_p に対して, 完全な A ハッシング関数の存在を決定するには, $O(n \log n)$ の時間がかかる. これは, 文字の取り出し (n), ソート ($n \log n$), そして比較 (n) に要する時間である. i_1, \dots, i_p の選び方は $2^l - 1$ 通りで, つまり全体として, 与えられたキーの集合に対して完全な A ハッシング関数の存在を決定するには, $O(2^l \cdot n \log n)$ となる. しかし

通常, l は 10 程度より小さいので, 2^l は定数と考えると, n の多項式時間で決定できるといえる.

2.3. 格納率について

定理 1 の証明で示したような連想値の割り当て方では, 完全ハッシング関数の格納率がかなり低下すると予想される. この方法と, Cichelli²⁾ が示したようなバックトラッキングによる連想値の割り当ての方法で, 格納率を比較してみる.

いま, m 個の異なる文字から重複を許して p 個を取り出すすべての組み合わせを表わす文字列の集合 I を考え (例 4), これを m - p 飽和集合と呼ぶ.

〔例 4〕

3—3 飽和集合 I は, $\Sigma = \{A, B, C\}$ とするとき $I = \{AAA, AAB, ABB, BBB, AAC, ABC, BBC, ACC, BCC, CCC\}$ である.

〔注意 3〕

m - p 飽和集合 I の要素の数 n は

$$n = {}_m H_p = \frac{(m+p-1)!}{p!(m-1)!}$$

である.

m - p 飽和集合 I に対して, バックトラッキングによって決定される連想値の割り当ては,

$$a(r_j) = \frac{p^{j-1} - 1}{p-1} + C$$

($1 < p, 1 \leq j \leq m, C$ は任意の整数) である. このことは, 次のようにして説明できる. $a(r_1) < a(r_2) < \dots < a(r_m)$ と仮定する. 実際, I の, 文字に対する対称性から, $i \neq j$ ならば $a(r_i) \neq a(r_j)$ でなければならない.

いま, $a(r_1) = C$ とする. 次に, r_1 と r_2 だけからなるキーに対して, 完全性を保ち, かつハッシュテーブルを可能な限り小さくするために $a(r_2) = C+1$ を割り当てる. ここで, r_1 と r_2 だけからなるキーの中で, ハッシュ値が最も大きいものは $(C+1) \times p$ となる. さらに, r_1, r_2, r_3 だけからなるキーに対して, 完全性を保ち, かつハッシュテーブルを可能な限り小さくするために $a(r_3) = C+1+p$ を割り当てる. こうして $a(r_i) = C+x$ が割り当てられれば $a(r_{i+1}) = C+x+p^{i-1}$ が割り当てられる (例 5).

〔例 5〕

3—3 飽和集合に対するバックトラッキングによる連想値の割り当ては, 次のようになる.

Function: $h(w) = a(w^{(1)}) + a(w^{(2)}) + a(w^{(3)})$
 Associated Values: $a(A) = 0, a(B) = 1, a(C) = 4$

I (Set of keys)	Hash Values
$w_1 = AAA$	$h(w_1) = 0$
$w_2 = AAB$	$h(w_2) = 1$
$w_3 = ABB$	$h(w_3) = 2$
$w_4 = BBB$	$h(w_4) = 3$
$w_5 = AAC$	$h(w_5) = 4$
$w_6 = ABC$	$h(w_6) = 5$
$w_7 = BBC$	$h(w_7) = 6$
$w_8 = ACC$	$h(w_8) = 8$
$w_9 = BCC$	$h(w_9) = 9$
$w_{10} = CCC$	$h(w_{10}) = 12.$

さて、定理1の証明で示した連想値の割り当て $a(r_j) = (p+1)^{j-1}$ を用いた $m-p$ 飽和集合に対する A 完全ハッシング関数の格納率 lf は、

$$lf = \frac{{}_m H_p}{(p+1)^m}$$

となる。一方、バケットラッキングによる割り当て $a(r_j) = (p^{j-1}-1)/(p-1)$ を用いた場合の格納率は、

$$lf = \frac{{}_m H_p}{(p^m-1)/(p-1)}$$

となり、飽和集合のような性質の悪い集合に対する A 完全ハッシング関数の場合では、連想値の割り当て方による格納率の差には、オーダーから見ても大差がない。

2.4. 最小完全な関数の決定可能性

完全な A ハッシング関数は存在するが、最小完全な A ハッシング関数は存在し得ないようなキーの集合 I がある (例6)。

[例6]

キーの集合として、 $I = \{w_1, \dots, w_n\}$ を考える。ここに、 $2 \leq n, w_k = r_k r_{k+1} (1 \leq k \leq n-1), w_n = r_1 r_n$ とする。

Function: $h(w) = a(w^{(1)}) + a(w^{(2)})$

Ex. 1 ($n=2$)

- $w_1 = AB$
- $w_2 = BA$

Ex. 2 ($n=3$)

- $w_1 = AB \quad h(w_1) = 1 \quad a(A) = 0, a(B) = 1,$

- $w_2 = BC \quad h(w_2) = 3 \quad a(C) = 2$
- $w_3 = AC \quad h(w_3) = 2$

Ex. 3 ($n=4$)

- $w_1 = AB \quad h(w_1) = 0 \quad a(A) = 0, a(B) = 0,$
- $w_2 = BC \quad h(w_2) = 1 \quad a(C) = 1, a(D) = 2$
- $w_3 = CD \quad h(w_3) = 3$
- $w_4 = AD \quad h(w_4) = 2$

Ex. 4 ($n=5$)

- $w_1 = AB \quad h(w_1) = 0 \quad a(A) = 0, a(B) = 0,$
- $w_2 = BC \quad h(w_2) = 1 \quad a(C) = 1, a(D) = 2,$
- $w_3 = CD \quad h(w_3) = 3 \quad a(E) = 2$
- $w_4 = DE \quad h(w_4) = 4$
- $w_5 = AE \quad h(w_5) = 2$

Ex. 5 ($n=6$)

- $w_1 = AB$
- $w_2 = BC$
- $w_3 = CD$
- $w_4 = DE$
- $w_5 = EF$
- $w_6 = AF$

Ex. 6 ($n=7$)

- $w_1 = AB \quad h(w_1) = 1 \quad a(A) = 0, a(B) = 1,$
- $w_2 = BC \quad h(w_2) = 2 \quad a(C) = 1, a(D) = 2,$
- $w_3 = CD \quad h(w_3) = 3 \quad a(E) = 3, a(F) = 3,$
- $w_4 = DE \quad h(w_4) = 5 \quad a(G) = 4$
- $w_5 = EF \quad h(w_5) = 6$
- $w_6 = FG \quad h(w_6) = 7$
- $w_7 = AG \quad h(w_7) = 4$

[注意4]

例6のようなキーの集合 I で、 $n = 4n' + 2$ (n' は0以上の整数) のとき、最小完全な A ハッシング関数は存在しない。

(証明)

$n = 4n' + 2$ のとき、最小完全な A ハッシング関数が存在するなら、

$$h(w_1) + h(w_2) + \dots + h(w_{4n'+2}) = b + (b+1) + \dots + (b+4n'+1) = 2\{(2n'+1)b + (4n'^2 + 3n')\} + 1 \text{ (奇数)}$$

となる。一方、完全な A ハッシング関数は、 $p = 2, i_1 = 1, i_2 = 2$ の場合だけにしか存在しないことは明らかなので、

$$h(w_1)+h(w_2)+\cdots+h(w_{4n'+2})=2\{a(A)+a(B) \\ +\cdots+a(r_n)\} \quad (\text{偶数})$$

となり、矛盾が生じる。

従って、背理法により、 $n=4n'+2$ のときには最小完全な A ハッシング関数が存在しないことが証明された□

次に、最小完全な A ハッシング関数の存在性に関する決定問題について考えよう。いま、

$$\mu_{kh} = |w_k|_{r_h}^{i_1, \dots, i_p}$$

とすると、ある p, i_1, \dots, i_p で、最小完全な A ハッシング関数の存在は、

$$A \times x = b \quad (1)$$

なる x の存在と同値である。ここに、

$$A = \begin{pmatrix} \mu_{11} & \mu_{12} & \cdots & \mu_{1m} \\ \mu_{21} & \mu_{22} & \cdots & \mu_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{n1} & \mu_{n2} & \cdots & \mu_{nm} \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix}, \quad b = \begin{pmatrix} b \\ b+1 \\ \vdots \\ b+n-1 \end{pmatrix}$$

とし、 x_1, \dots, x_m は整数、 b は 0 以上の整数で、 b はその要素の置換を許すものとする。また、 A に関して、

$$\begin{aligned} \mu_{11} + \mu_{12} + \cdots + \mu_{1m} &= \mu_{21} + \mu_{22} + \cdots + \mu_{2m} \\ &= \cdots \\ &= \mu_{n1} + \mu_{n2} + \cdots + \mu_{nm} = p \end{aligned}$$

なる関係がある。そこで、もしこの方程式(1)の解 x_1, \dots, x_m 、および b が 1 組見つければ、 $x'_1 = x_1 + 1, x'_2 = x_2 + 1, \dots, x'_m = x_m + 1, b' = b + p$ として、やはりこの方程式の別の解 x'_1, \dots, x'_m 、および b' の組を作ることが可能で、従って無限個の解の組が存在することになる。このことから、もし、この方程式の解が存在するならば、 $0 \leq b < p$ となる解が必ず存在する。つまり、 b は有限個について調べればよい。

さて、(1)が解(有理数解)を持つための必要十分条件は、

$$\text{rank} A = \text{rank}(Ab) \quad (2)$$

である。従って、この条件(2)は(1)が整数解を持つための必要条件である。この必要条件(2)を満たす

$$\text{rank} A = \text{rank}(Ab) = r$$

について、

i) $r \leq m$ のとき、解(有理数解)は一意に決

まり、それが整数解であるか否かは決定可能である。

ii) $r > m$ のとき、(1)は連立一次不定方程式となる。1つの一次不定方程式 $a_1 x_1 + a_2 x_2 + \cdots + a_m x_m = b$ が整数解を持つための必要十分条件は、

$$(a_1, a_2, \dots, a_m) | b \quad (3)$$

である。ここに、 (a_1, a_2, \dots, a_m) は a_1, a_2, \dots, a_m の最小公倍数である。(1)は、 n 個の一次不定方程式より成っていて、これを次のように順次解いていく。

まず第 1 の一次不定方程式が条件(3)を満たしていれば、任意の整数のパラメーターを含む一般解を得る。この一般解を第 2 の一次不定方程式に代入して、同様に一般解を求め、これを次々に第 n の一次不定方程式にまで繰り返せば、(1)の、整数の一般解が得られる⁵⁾。もし、この繰り返しの途中で、条件(3)を満たさない一次不定方程式があれば(1)に整数解は存在しない。

以上のことから、次の結果が得られる。

[定理 2]

キーの集合 I に対して、最小完全な A ハッシング関数が存在するか否かを問う問題は、決定可能である。

(証明)

長さ l のキーから何番目の文字を取り出すかという組み合わせは $2^l - 1$ 通りで、有限である。その 1 つ 1 つの場合に対して、 b における b は p 通り、また置換は $n!$ 通りで、いずれも有限である。

こうして各要素が具体的に決まった方程式(1)については、整数解の存在性が可解であったので、上記のすべての場合をつくせば、キーの集合 I に対して最小完全な A ハッシング関数が存在するか否かという問題は決定可能である□

3. Cichelli の完全ハッシング関数

本節では、Cichelli が定義したハッシング関数について、§ 2 と同様の議論が可能であることを述べ、与えられたキーの集合 I に対して Cichelli の完全ハッシング関数が存在するための必要十分条件を示す。また、Cichelli の関数と A ハッシング関数の関係について考察する。

3.1. 定義

本節では、キーの集合として、 $I = \{w_1, w_2, \dots, w_n\} \subset \Sigma^+$ を考える。すなわち、同じキーの集合の中で、

必ずしもキーの長さは等しくないものとする。今後、キー $w \in I$ の長さを $l(w)$ と表わすことにする。

〔定義8〕

a を連想とすると、

$$h(w) = l(w) + a(w^{(1)}) + a(w^{(l(w))}) \quad (w \in I)$$

で定義される関数 $h: I \rightarrow N$ を Cichelli のハッシング関数という。

〔例7〕

キーの集合として、出現頻度の高い英単語を考える²⁾。

Function: $h(w) = l(w) + a(w^{(1)}) + a(w^{(l(w))})$

Associated Values:

$a(A) = 3$	$a(B) = 15$	$a(C) = 0$	$a(D) = 7$
$a(E) = 0$	$a(F) = 15$	$a(G) = 0$	$a(H) = 10$
$a(I) = 0$	$a(J) = 0$	$a(K) = 0$	$a(L) = 0$
$a(M) = 12$	$a(N) = 13$	$a(O) = 7$	$a(P) = 0$
$a(Q) = 0$	$a(R) = 12$	$a(S) = 6$	$a(T) = 0$
$a(U) = 15$	$a(V) = 0$	$a(W) = 14$	$a(X) = 0$
$a(Y) = 9$	$a(Z) = 0$		

I (Set of keys)

Hash Values

$w_1 = A$	$h(w_1) = 7$
$w_2 = AND$	$h(w_2) = 13$
$w_3 = ARE$	$h(w_3) = 6$
$w_4 = AS$	$h(w_4) = 11$
$w_5 = AT$	$h(w_5) = 5$
$w_6 = BE$	$h(w_6) = 17$
$w_7 = BUT$	$h(w_7) = 18$
$w_8 = BY$	$h(w_8) = 26$
$w_9 = FOR$	$h(w_9) = 30$
$w_{10} = FROM$	$h(w_{10}) = 31$
$w_{11} = HAD$	$h(w_{11}) = 20$
$w_{12} = HAVE$	$h(w_{12}) = 14$
$w_{13} = HE$	$h(w_{13}) = 12$
$w_{14} = HER$	$h(w_{14}) = 25$
$w_{15} = HIS$	$h(w_{15}) = 19$
$w_{16} = I$	$h(w_{16}) = 1$
$w_{17} = IN$	$h(w_{17}) = 15$
$w_{18} = IS$	$h(w_{18}) = 8$
$w_{19} = IT$	$h(w_{19}) = 2$
$w_{20} = NOT$	$h(w_{20}) = 16$
$w_{21} = OF$	$h(w_{21}) = 24$
$w_{22} = ON$	$h(w_{22}) = 22$

$w_{23} = OR$	$h(w_{23}) = 21$
$w_{24} = THAT$	$h(w_{24}) = 4$
$w_{25} = THE$	$h(w_{25}) = 3$
$w_{26} = THIS$	$h(w_{26}) = 10$
$w_{27} = TO$	$h(w_{27}) = 9$
$w_{28} = WAS$	$h(w_{28}) = 23$
$w_{29} = WHICH$	$h(w_{29}) = 29$
$w_{30} = WITH$	$h(w_{30}) = 28$
$w_{31} = YOU$	$h(w_{31}) = 27$

この場合は最小完全なハッシング関数になっている。

3.2. $l(w)$ の扱い方に関する注意

キーの集合 I の要素のうち、最長のキーの長さを

$$L = \max l(I)$$

と表わす。いま、1 から L まで L 通りのキーの長さに関する情報を L 個の異なる文字の情報に置き換えて、連想値を $a(l(w)) = l(w)$ と割り当てていると考えることが可能である。こうすると § 2 とほぼ同様の議論が可能となる。

さらに、 $a(l(w))$ を自由に割り当てることを許せば、§ 2 における $p = 3, m = m + L$ の場合とまったく同様の議論が可能となるが、Cichelli の方法ではこれは許されていない。

3.3. 完全な関数の存在条件

$l(w)$ を 3.2 で述べたように扱おうと、§ 2 とほぼ同様の議論で完全な Cichelli のハッシング関数が存在するための必要十分条件を与えることが可能である。

〔定理3〕

キーの集合 I に対して、Cichelli の完全ハッシング関数が存在する。

\Leftrightarrow 任意の $u, v \in I, u \neq v$ に対して、

- i) $l(u) \neq l(v)$ または、
- ii) $\{u^{(1)}, u^{(l(u))}\} \neq \{v^{(1)}, v^{(l(v))}\}$.

(証明)

\Leftarrow $a(r_j) = (L+2)^j (1 \leq j \leq m)$ とすると、

$$h(w) = l(w) + |w|_{j_1}^{(1, l(w))} \times (L+2) + \dots + |w|_{j_m}^{(1, l(w))} \times (L+2)^m$$

となり、 $1 \leq l(w) \leq L, 0 \leq |w|_{j_j}^{(1, l(w))} \leq 2$ より、 $h(w)$ は $(m+1)$ 桁の $(L+2)$ 進数として表わされる。しかも、

- i) $l(u) \neq l(v)$ または、
- ii) $\{u^{(1)}, u^{(l(u))}\} \neq \{v^{(1)}, v^{(l(v))}\}$

より、任意の、 I の異なる2要素 u, v に対して、

$h(u) \equiv h(v)$ となる.

⇒) Cichelli の完全ハッシング関数が存在する.

⇔ $\exists a(r_1), \dots, a(r_m): \forall u, v \in I, u \neq v$ に対して

i) $h(w) = l(w) + a(w^{(1)}) + a(w^{(2)})$

ii) $h(u) \equiv h(v)$

より明らか□

[注意5]

キーの集合 I に対して Cichelli の完全ハッシング関数の存在を決定するには、 $O(n \log n)$ の時間がかかる.

3.4. Cichelli の関数と A 関数の関係

キーの集合 I のすべての要素の長さが l に等しい場合には、Cichelli の関数は

$$h(w) = l + a(w^{(1)}) + a(w^{(2)})$$

となる. これは、 $p = 2, i_1 = 1, i_2 = l$ としたときの A ハッシング関数と、最小性、完全性の点で等価である.

さらに、§ 2 で A ハッシング関数ではキーの長さがすべて l に等しいと仮定したが、ここでこの仮定を取り除き、 $l(w) < i$ のとき $a(w^{(i)}) \equiv a(_)$ と定義する. この定義は、§ 2 で議論した内容と何ら矛盾はせず、さらに $f_i(w)$ をキー $w \in I$ の持つ固有の情報としたとき、 $f_i(w) \in \Sigma$ と考えて、

$$h(w) = a(f_1(w)) + a(f_2(w)) + \dots + a(f_i(w)) \quad (w \in I)$$

で定義される関数を A' ハッシング関数と呼ぶことにする. A' ハッシング関数は A ハッシング関数を拡張解釈したものと考えることができる.

[定理4]

Cichelli の関数は A' ハッシング関数である.

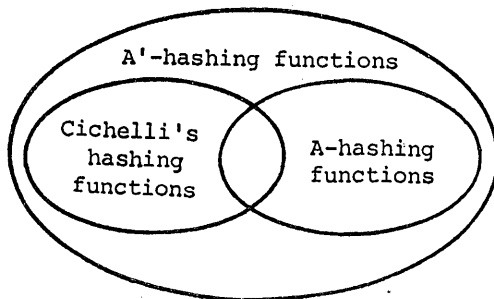


Fig. 1 Relation of 3 types of hashing functions.

(証明)

Cichelli の関数は、

$$h(w) = a(l(w)) + a(w^{(1)}) + a(w^{(2)})$$

と表わすことができる. ここで、 $l(w)$ は § 3.2 で述べたように一種の文字と見なして、 $a(l(w))$ の値は、 $l(w)$ であるとする□

以上のことから、 A ハッシング関数、 A' ハッシング関数、Cichelli の関数の族としての関係は Fig. 1 のベン図で表わされる.

4. 関数値の計算に要するコスト

完全な A ハッシング関数を用いて、ハッシュテーブルの中から1個のキーを探索するための時間的コスト T_A は次のように定義するのが妥当である.

$$T_A = p(t_1 + t_a) + t_2 + t_c$$

また、これと比較するため、Jaeschke⁶⁾ が定義した最小完全ハッシング関数

$$h(w) = [C/(Dw + E)] \bmod n$$

の場合に要する時間的コスト T_J を

$$T_J = t_m + t_a + t_d + t_3 + t_2 + t_c$$

と定義する. ここに、

p : 連想値を取り出す文字の個数

t_a : 整数の足し算1回に要する時間

t_m : 整数の掛け算1回に要する時間

t_d : 整数の割り算1回に要する時間

t_c : キー比較1回に要する時間

t_1 : 1つの文字の連想値を知るのに要する時間

t_2 : ハッシュテーブルからキーを1つ呼ぶのに要する時間

t_3 : 整数の剰余を1回求めるのに要する時間

C, D, E : 整数

とする. 九州大学大型計算機センターの FACOM M-200 における FORTRAN-GE によるプログラム実験では、 $t_a = 1$ としたとき、 $t_m = 1, t_d = 20, t_1 = 1, t_3 = 22$ であった. これらを $T_A < T_J$ の式に代入すると、

$$p < 22 \text{ のとき } T_A < T_J$$

となる. 通常、 p は 1~5 程度であるから、 A 完全ハッシング関数の方が十分に速いといえる.

また、同様の比較を Dividing 法⁷⁾ によるハッシング関数に対しても行なったが、 $p < 11.5$ のとき、 A

完全ハッシング関数の方が速いという結果を得た。

5. おわりに

静的なキーの集合に対する A 完全ハッシング関数の特徴を示し, Cichelli が定義した完全ハッシング関数が存在するための必要十分条件を与えた。また, 関数値を計算するための時間的コストの点で, 他のハッシング関数と比較して, A 完全ハッシング関数が優れていることを示した。

今後は, 最小完全な関数の存在するキーの集合に対して, その連想値を割り当てる方法, および, A 完全ハッシング関数が存在するか否かを決定する問題の複雑性 (キーの長さ l に対して NP 完全であることが予想される) 等について検討していきたい。

参 考 文 献

1) Sprugnoli, R. : Perfect Hashing Functions: A Single Probe Retrieving Method

for Static Sets, CACM, 20, 11, (Nov. 1977).
 2) Cichelli, R. J. : Minimal Perfect Hash Functions Made Simple, CACM, 23, 1, (Jan. 1980).
 3) Jaeschke, G. and Osterburg, G. : On Cichelli's Minimal Perfect Hash Functions Method, CACM, 23, 12, (Dec. 1980).
 4) 玉越, 有川: 連想値を用いた完全ハッシング関数, 昭和56年度電気四学会九州支部連合大会講演論文集.
 5) 高木貞治: 初等整数論講義第2版, 共立出版.
 6) Jaeschke, G. : Reciprocal Hashing: A Method for Generating Minimal Perfect Hashing Functions, CACM, 24, 12, (Dec. 1981).
 7) Martin, J. : Computer Data-Base Organization, 2nd edition, Prentice-Hall, Inc., 1977.
 8) 玉越, 有川: 連想値を用いる完全ハッシング関数の存在条件について, 京都大学数理解析研究所講究録, 1982.