數學與資料存取

資管系一年級 張靖群

一、介紹

孫武是我國歷史上一位非常出名的將軍，在吳王闔廬殿前擔任大將，用兵如神，曾助吳王，北威齊晉，西破強楚，建立了許多豐功偉業。他也是我國歷史上鮮有的偉大數學家，著有孫子算經。然而卻很少人能想像得到這位已經作古兩千五百年的將軍所提出的孫子定理卻可以被用來做資料庫的資料存取。

在孫子算經中曾經提到一個數學問題「物不知其數」：「今有物不知其數，三三數之剩二，五五數之剩三，七七數之剩二，問物幾何？」亦即求正整數 C，使得

C≡2 (mod 3)，
C≡3 (mod 5)，
C≡2 (mod 7)。

然而，C 值是否必定存在？在孫子算經中亦提到了一個定理，名為孫子定理，又叫中國餘數定理。

任予 2N 個正整數 $m_1, m_2, \cdots, m_N$和 $r_1, r_2, \cdots, r_N$。則存在一個常數 C，使得

C≡r_1 (mod m_1)，
C≡r_2 (mod m_2)，
…
C≡r_N (mod m_N)，其中對於 $i \neq j$，$m_i \neq m_j$。

由孫子定理，我們知道：令$m_1 = 3$，$m_2 = 5$，$m_3 = 7$，且令$r_1 = 2$，$r_2 = 3$，$r_3 = 2$，則必定可以找到一個整數 C，令此數為 23，滿足

C mod $m_1 = 23$ mod 3 $= 2$，
C mod $m_2 = 23$ mod 5 $= 3$，
C mod $m_3 = 23$ mod 7 $= 2$。

古代對此問題的實際解法，有一歌訣：

「三人同行七十稀，
五樹梅花廿一枝，

七子團圓正半月，
除百零五便得知。」

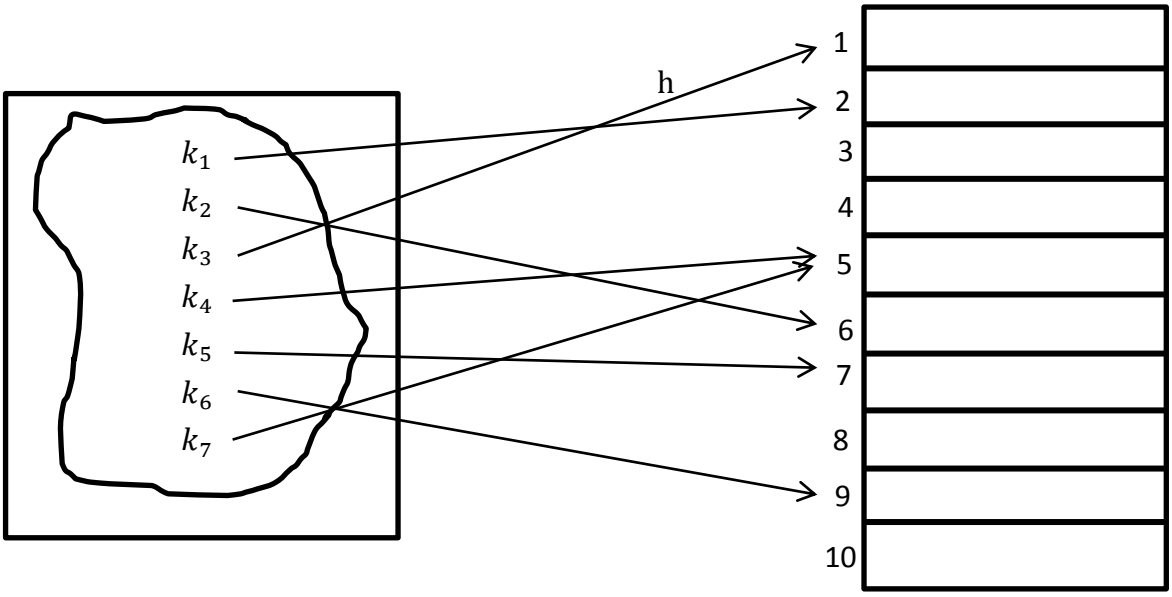
意為：以 70 乘用 3 除所得之餘數加上以 21 乘用 5 除所得之餘數再加上以 15 乘用 7 除所得之餘數，然後用 105 之倍數加減之；因此解法為 2 x 70 + 3 x 21 + 2 x 15 = 233，減去 105 的二倍，得 23，此乃所求之數。

二、資料存取

由於資訊時代的來臨，應用電腦做為資料處理、管理決策分析、航空訂位系統、電腦 e 化教學、國防應用、犯罪鑑識工具越趨普及。而廣泛採用電腦，用以存取資料遂成為電腦應用的主要動機。然而，如何設計一套有效的資料存取技術以提高電腦的運作效率是電腦科學家們非常感興趣的研究主題。

赫序 (Hashing) 是一種相當有效的資料存取技術，該技術被廣泛地應用在設計資料庫管理系統。利用赫序函數，我們可以將每一個關鍵詞對應到一個整數，而該數值可以看成關鍵詞所存放的位址。

舉例而言，假設我們有 7 個關鍵詞，分別是七個同學的學號，稱之為k_1, k_2, k_3, k_4, k_5, k_6 和 k_7，令 S = {1, 2, 3, 4, 5, 6, 7, 8, 9, 10} 為一組有 10 個位址的位址空間。下圖表示關鍵詞集合至位址空間的對應關係，而此對應函數 h 稱作赫序函數。



由上圖，我們發現k_4 和 k_7 均對應到 5，當有二個或多於兩個關鍵詞同時對應到一個位址時，我們稱此現象為「碰撞」。此時在圖中，k_4 和 k_7 經過赫序函數 h 的映射，卻撞在一起。過去有許多專家曾經研究如何解決「碰撞」所帶來的資料擠不下一個位址的問題，我們稱這一派的學者為「亡羊補牢派」。

假使資料經過赫序函數的轉換，沒有出現任何碰撞的現象，亦即關鍵詞集合和函數空間為一對一之對應，則我們稱此函數為完美赫序函數。這一派的學者經

常被稱為「未雨綢繆派」。顯然欲解決碰撞問題最好的辦法是建造一個完美赫序函數，一旦我們可以建造成功關鍵詞和位址空間的一對一關係，資料的存取就十分便捷了。

一般而言，要建造一個讓關鍵詞的映射完全沒有碰撞的完美赫序函數是一件很困難的事情。在 1984 年，Chang 利用了孫子定理的性質來設計一個完美赫序函數。讓我們用一個例子來說明 Chang 的方法。令張三、李四、王五、趙六等四位同學的學號分別為 4, 5, 7, 9。且令記憶體共有四個位址。分別為位址 1， 位址 2， 位址 3，位址 4。則 Chang 所設計的完美赫序函數：

$h(m_i) = \mathrm{C} \mod m_i$

$h(m_i) = 157 \mod m_i$, 其中m_i為第i位同學的學號，$h(m_i)$為映射位址。

亦即，

令$m_1 = 4$，$m_2 = 5$，$m_3 = 7$，$m_4 = 9$，且令$r_1 = 1$，$r_2 = 2$，$r_3 = 3$，$r_4 = 4$，則 $\mathrm{C} = 157$ 滿足

$\mathrm{C} \mod m_1 = 157 \mod 4 = 1$，

$\mathrm{C} \mod m_2 = 157 \mod 5 = 2$，

$\mathrm{C} \mod m_3 = 157 \mod 7 = 3$，

$\mathrm{C} \mod m_4 = 157 \mod 9 = 4$。

由孫子定理的充分條件知道：倘若所給予的一組關鍵詞 $\{m_1, m_2, \ldots, m_N\}$ 滿足任兩個關鍵詞均彼此互質的話，我們可以令 $m_1 < m_2 < \ldots < m_N$，且$r_i = \mathrm{i}$, i =1, 2, … , N，則一定存在一個常數C，使得函數$h(m_i) = \mathrm{C} \mod m_i = \mathrm{i}$ 為一個完美赫序函數。

Chang [1984] 亦提出了一個計算赫序函數中唯一常數 C 值的方法，他的解法概述如下：

(1)計算 $M_i = m_1 \times m_2 \times \ldots \times m_{i-1} \times m_i \times \ldots \times m_N$。

(2)計算 b_i，滿足$M_i b_i \equiv 1 \pmod{i}$，i=1, 2 , … , n。

(3)計算 C= $\sum_{\mathrm{i}=1}^{\mathrm{n}} M_i\, b_i$ i。

然而問題是如果這些關鍵詞m_1, m_2, … , m_N不滿足兩兩互質時，怎麼辦呢?

給予一個有限的正整數集合 K= $\{m_1, m_2, \ldots, m_N\}$。如果 P($m_i$)為一個定義在集合 K 的質數產生函數，則必定存在一個常數 C，使得$h(m_i) = \mathrm{C} \bmod \mathrm{P}(m_i)$ 為一最佳完美赫序函數。

在數論導引[1975]一書中引述了 5 個存在質數產生函數：

(1) P(x) = $x^2 - x + 17$, 1≦x≦16，

(2) P(x) = $x^2 - x + 41$, 1≦x≦40，

(3) P(x) = $x^2 + x + 41$, 1≦x≦39，

(4) P(x) = $x^2 - 81x + 1681$, 41≦x≦80，

(5) P(x) = $x^2 - 79x + 1601$, 40≦x≦79。

將來是否可以找到更多可以產生更多質數的質數產生函數? 將是一個非常有趣的數學問題。

三、結論

數學的發展迄今已有數千年以上的歷史了，其優美的結構，豐富的內涵常讓我們身為後代子孫的研習者大為佩服。從孫子定理導引出一套赫序函數的設計方法，配合質數產生函數，可以應用到計算機資料庫裡的快速資料存取。將來如何將更多有用的數學定理和精彩的數學理論導引到管理與工程的實務應用，將是我們學習應用科學的人應該努力的研究目標。

參考文獻

1. Chang, C. C., (1984): The Study of an Ordered Minimal Perfect Hashing Scheme, *Communications of the Association for Computing Machinery*, Vol.27, No. 4, April 1984, pp. 384-387.

2. 華羅庚, (1975): 數論導引, 先登出版社, 1975.