行政院國家科學委員會專題研究計畫成果報告

代數與離散結構之解析與隨機性質之研究之

子計畫四: 應用性學科中組合結構之計算性研究(3)

Computational Studies of Combinatorial Structures in Various Applied Disciplines, Part III

計畫編號: NSC - 89 - 2115 - M - 032 - 002
執行期限: 1999 年 8 月 1 日至 2000 年 7 月 31 日

計畫總主持人: 葉永南　　中央研究院數學所
子計畫主持人: 楊柏因　　淡江大學數學系

1 中文摘要

沿襲在前二年的腳步, 我們賡續對幾個常見的演算法, 特別是排序的應用, 做出有實用性的改進方式, 並做了一些以前沒有的漸近行爲分析。同時, 我們進行一些其他的組合結構生成函數演算, 把一些原有計數性組合的結果推廣。結果並簡述於下。

關鍵字: 演算法, 化學數學, 生成函數, 組合結構

Abstract

We followed the footsteps of the previous years of execution, arriving at some improvements of algorithms that has practical value. For some algorithms, we arrived at asymptotic descriptions of behaviors heretofore unknown. We also extended some enumerative combinatorial results to their q-analog or generating function forms.

Keywords: Algorithms, Chemical Mathematics, combinatorial structures, generating functions

2 緣由與目的

在準備展開整個整合型計畫的工作時, 我們曾發表過這樣的看法-- 當時我們認爲, 計算機可以幫助組合學家獲得很多新的方向與內涵:

計算機現已廣泛而深遠的被應用於各門學科, 也刺激了近年來離散結構的蓬勃研究與發展。這個趨勢不僅引進了大量新的結構, 同時也給既有的理論開啓了許多新的研究方向。

另一方面傳統的組合學研究在許多的數學家努力耕耘下, 現在也已進入一個更有系統與更抽象層次的階段。例如：在計數組合學（Enumerative Combinatorics）中就已引入了大量代數技巧, 並加以純化萃煉, 使其更加獨立、完整、且更能有系統的在其他學科應用。

在這樣的時代潮流之下, 代數性質的研究逐漸完備再次刺激其他性質, 例如漸近 (asymptotic) 與隨機 (probabilistic) 性質的進一步探討。這兩類帶有高度相關的研究對於離散結構而言可謂不可或缺 (很可惜地, 卻爲大多數的組合學家所忽略), 其主要原因之一是: 大多數的計數組合問題都有所謂的組合爆炸現象, 再加上大多數的計算公式都十分複雜, 雖可借助於如 Lagrange Inversion, 或 Wilf-Zeilberger Summation Method 的技巧, 或電腦代數軟體來幫助, 但是在實用上還是經常有必要尋求漸近或逼近式, 而樣本空間過於龐大, 也使得隨機性質的探討成爲必然之勢。

事實上, 今天這些目標依然有效, 但是與幾年前不同的, 根據 Moore's Law 繼續在日新又新的電腦技術, 使得我們有更多的東西可以模擬演練, 高度的平行化處理更是很多原本僅爲純理論性的技術現在可以有效的應用在實際上。

3 結果

我們分爲兩大類問題的探討來敘述:

3.1 計算機科學上的探討

結果尚在整理中, 摘要如下:

D. Knuth 曾說電腦有四分之一的時間花在處理排序上，因之排序是一個典型並且重要性毋庸置疑的演算法。據說這個數字到今天仍然不變。排序的一種演算法, Hoare 氏的 **Quicksort**– 也就是在陣列中取樣做爲 Key Value, 並以之把原陣列分割爲較 Key 高與低的兩段, 的遞迴式運作– 也因之變成所有現代操作系統的系統內建工具副程式 (system library subroutine) 的一部分。但是 Quicksort 在兩個方面都有加強的空間:

- 它並不穩定, 也就是說在最壞狀況所花的時間遠大於平均時間, 而接近最壞狀況的結果又經常在現實生活所可能引入的輸入值中發生。
- 在分割上可以採取比較均勻的方式, 但是想要分割的越平均, 就必須花更多的前置作業時間來達成這個目標。

這是我們在[22]中所做的觀察:

> 由於 Quicksort 自 D. Knuth 與 R. Sedgwick 以來也是最爲人所分析的演算法之一, 我們迅速的達到一個結論: 所有的在基本方法上所做的變形或是改進，例如 samplesort, 或是更近期的 leapfrogging samplesort, 都浪費到太多前置作業的時間; 對於一千萬筆資料的大資料庫, 或許它們值得採用, 但是對於一個經常使用的系統副程式而言, 它們是接受到“報酬遞減律”的制約, 因此是不經濟的。

去年, 我們按照“如何經濟的增加已排序度” 的觀念來排序, 分析出常見的取五樣本中位數, 七樣本中位數, 或更多的 $(2t+1)$樣本中位數等等都是相對不經濟的; 並深入探討了兩個變式: 一個是由 Tukey 提出的準九樣本中位數 (Ninther quicksort, 即取三組三樣本中位數之中位數) 分割法, 另一個是七樣本全排序法(7-Samplesort), (Green 文[12] 的衍伸)。在一些常見的現實參數下,我們對各種排序法的平均比較次數得到如下的表示:

$$
\begin{aligned}
Q(n) &\approx 2\, n \ln n - 2.84n && (1)\\
Q_3(n) &\approx 1.72\, n \ln n - 2.24n && (2)\\
Q_5(n) &\approx 1.62\, n \ln n - 1.97n && (3)\\
Q_{7s}(n) &\approx 1.64\, n \ln n - 9.03n && (4)\\
Q_{9p}(n) &\approx 1.57\, n \ln n - 6.44n && (5)
\end{aligned}
$$

其中 Q, Q_3, Q_5, Q_{7s}, Q_{9p} 分別表示正常, 三中位數, 五中位數, 7-samplesort, Ninther 的平均比較次數理論值。

因而, 我們知道 Q_{7s}, Q_{9p} 是優越的計算方式。但是現在我們更進一步的推展, 把 Q_{7s}, Q_{9p} 的想法揉合, 做出 Q_{3t} 這個排序法。這個做法的主要想法如下:

1. 先假設要排序的東西數 $n = 3^k$, 如否, 我們可適當調節原輸入使得到相同的結果。
2. 三三一組做中位數, 並遞迴的如此做, 以此做出一個中位數的塔狀結構。
3. 以頂段的超中位數 (supermedian, 或remedian 見[17]) 爲中心做quicksort 的分割。
4. 繼續用下階的中位數繼續分割兩端，直到小到某一 n 大小時, 改爲insertion 或heapsort。

可證本超中位塔製造時間(比較次數)約 $4n/3$, 我們可以得到一個上限即可以達成Q_{3t} 的比較次數小於$3n \ln n/2 - cn$, $c \approx 5$ (理論上的最佳係數是 $n \log_2 n = 1.44n \ln n$), 但如要造成嚴格的sort-in-place 則將稍較慢, 我們正在探討其可能的變形。但是我們可以斷定這是有實際效果的演算法。

我們找到一些細部的其他演算法改良，不過我們感到興趣的是可以證明

計算超過五個以上數字的 median 以做Quicksort 排序總是不如 remedian 造塔方法。

3.2 其他

我們對很多其他的學科作了一些計算性工作，結果尚在整理中。在我們已經有所研究的 Wiener Polynomial (即圖形中頂點間距離的生成函數) 上找到了一些非遞迴性的公式可以計算化學上有意義的圖Crown Graph 的 Wiener Polynomial 也對 partition 和 composition 總體的漸近性質多作了一些探討。

目前我們最有興趣的結果來自於一個組合生成的問題，即在某些限制下的組合分配 (distributive combinations, 來自於橋戲理論的受限生成問題)，與 Uni. of Oregon 的 M. Ginsberg 共同找到最佳化的解法，可以用於改良現行的橋藝 AI, 現用於目前最強的橋藝程式 GIB 中，將整理後發表。

4 討論與計畫成果自評

我們最近的努力方向在使用組合學的概念於計算機領域，但引入分析學的技巧，希望藉由改換觀點來尋找一些問題的新處方 (prescription)。

這一兩年來幾個試驗中的例子包含前面所述的Quicksort 改良法，根部插入式(Root-Insertion) Quicksort construction, 如何判別各式sort 方法所生成的搜尋樹等等。最近隨著計算機的變快變強，我們有更多的素材可以研究

我們相信不論是理論或是實用性上來說，我們的結果都將是相當有用的，特別是一些排序法已經 submit 給世界上使用最多的系統副程式集 Glibc, 可能被採用於下一版中。

5 參考文獻

References

[1] G. E. Andrews, *The theory of partitions*, Encyclopedia of Mathematics and Its Applications, Volume 2, Edited by Gian-Carlo Rota, Addison-Wesley Publishing Company, 1976.

[2] E. A. Bender (1973). Central and local limit theorems applied to asymptotic enumeration. *Journal of Combinatorial Theory, Series A* **15** 91–111.

[3] E. A. Bender and L. B. Richmond (1983). Central and local limit theorems applied to asymptotic enumeration. II. Multivariate generating functions. *Journal of Combinatorial Theory, Series A* **34** 255–265.

[4] E. A. Bender, L. B. Richmond and S. G. Williamson (1983). Central and local limit theorems applied to asymptotic enumeration. III. Matrix recursions. *Journal of Combinatorial Theory, Series A* **35** 263–278.

[5] P. Erdős and M. Szalay, *On the statistical theory of partitions,* Colloquiua Mathematica Societatis János Bolyai, 34, Topics in classical number theory, Budapest, 397–450, 1981.

[6] P. Flajolet and A. M. Odlyzko, Singularity analysis of generating functions, *SIAM Journal on Discrete Mathematics*, 3, 216–240 (1990).

[7] P. Flajolet and M. Soria, Gaussian limiting distributions for the number of components in combinatorial structures, *Journal of Combinatorial Theory, Series A*, 53, 165–182 (1990).

[8] Foata, D. (1974). *La série génératrice exponentielle dans les problèmes d'énumération.* Les Presses de l'Université de Montréal, Montréal.

[9] B. Fristedt, The structure of random partitions of large integers, *Transactions of the American Mathematical Society*, **337** (1993), 703–735.

[10] W. M. Y. Goh and E. Schmutz, The number of distinct part sizes in a random integer partitions, *Journal of Combinatorial Theory, series A*, **69** (1995), 149–158.

[11] I. P. Goulden and D. M. Jackson, Combinatorial enumeration, John Wiley & Sons, New York, 1983.

[12] D. H. Green and D. E. Knuth, *Mathematics for the Analysys of Algorithms*

[13] J. C. Hansen, Order statistics for decomposable combinatorial structures, *Random Structures and Algorithms*, 5, 517–533 (1994).

[14] C. B. Haselgrove and H. N. V. Temperley, Asymptotic formulae in the theory of partitions, *Proceedings of the Cambridge Philosophical Society*, **50** (1954), 225–241.

[15] P. Henrici, *Applied and computational complex analysis*, three volumes, John Wiley & Sons, New York, 1977.

[16] J. Herzog, Weak asymptotic formulas for partitions free of small summands, II, *Acta Mathematica Hungarica*, **62** (1993), 173–188.

[17] *An Efficient Algorithm for the Approximate Median Selection Problem*, by S. Battiato, D. Cantone, D. Catalano, G. Cincotti and M. Hofri. Presented at the Fifth Seminar on the Analysis of Algorithms, June 1999, available at `http://www.cs.wpi.edu/ hofri/`, to be published.

[18] H.-K. Hwang, Distributions of integer partitions with large number of summands, *Acta Arithmetica*, accepted for publication.

[19] H.-K. Hwang, A Poisson $*$ geometric law for the number of components in unlabelled combinatorial structures, *Combinatorics, Probability and Computing*, accepted for publication.

[20] H.-K. Hwang, *Théorèmes limites pour les structures combinatoires et les fonctions arithmétiques*, Thèse, Ecole polytechnique, 1994.

[21] H.-K. Hwang and Y.-N. Yeh, Measures of distinctness for integer partitions and compositions, preprint, 1996.

[22] H.-K. Hwang, B.-Y. Yang and Y.-N. Yeh, *Presorting algorithms: an average-case point of view*, Theorectical Computer Science, (accepted, to appear).

[23] D. E. Knuth, *The art of computer programming, volume II, seminumerical algorithms*, Second edition, Addison Wesley, Reading, MA, 1981.

[24] R. P. Stanley, *Enumerative combinatorics*, Monterey, CA, Wadsworth, 1986.

[25] G. Szekeres, Asymptotic distributions of the number and size of parts in unequal partitions, *Bulletin of the Australian Mathematical Society*, **36** (1987), 89–97.

[26] R. Wong, *Asymptotic approximations of integrals*, Academic Press, Inc., Boston, 1989.

[27] B.-Y. Yang, and Y. Yeh, *Zigging and Zagging in Pentachains*, Advances in Applied Mathematics 16(1995) 72-94.