

運用進階基因演算法於選題策略之研究

A Study of Applying Advanced Genetic Algorithm to Select Test Items

孫光天

臺南師範學院資訊教育所
ktsun@ipx.ntntc.edu.tw

程千芬

臺南師範學院資訊教育所
tnccf7@mail.tn.edu.tw

蔡淑燕

臺南師範學院資訊教育所
tnnell95@mail.tn.edu.tw

摘要

近年來，在人工智慧領域上，基因演算法 (Genetic Algorithm) 對於解決組合最佳化問題已有不錯成果，然而在大量限制條件問題中，有效率的尋求一合理最佳解仍有待研究；本研究即以教育評量中，如何應用測驗訊息函數 (test information function) 選題，編製一份符合測驗設計者施測目的之測驗 (選題策略問題) 為例，改良基因演算法參數使用策略，使在大量限制條件下，仍能有效尋找出合理的最佳解或次佳解。因此，在本文中我們將提出“進階基因演算法”，以題目反應理論 (Item Response Theory, 簡稱 IRT) 為基礎，運用簡單基因演算法，再修改演化參數與目標函數應用技術，將其應用於有大量限制條件之選題問題。結果顯示本方法是一個高效率的方式，不但使測驗品質更符合測驗設計者需求，提高其實用性，且對教育評量研究者提供一更有效的工具，成功延伸人工智慧 - 基因演算法在教育測量領域之應用。

關鍵字：基因演算法，選題策略，題目反應理論，人工智慧，進階基因演算法

一、簡介

題目反應理論 (Item Response Theory, 簡稱 IRT) 是根據題目難度、鑑別度、猜測度及受試者能力指標來分析之測驗理論。在符合單向度條件下，存在下述幾點優勢：題目對受試者能力之估計是獨立的，題目難度和鑑別度與受試者取樣無關，能力估計之精確度是可估計的 [13]，且在測驗設計上足以應付不同的測驗目的 [3,6,11,14,23,29,30]。因此，題目反應理論在建構測驗，不僅有逐漸凌駕古典測驗理論，甚至有進而取代之趨勢 [1]。使用題目反應理論建構測驗的過程中，如何應用測驗訊息函數 (test information function) 選題，編製一份符合測驗設計者施測目的之測驗，在於選題策略的運用 [2]。而現今無論是國際性托福考試，或是國內基本學力測驗，都是由已經建立好的題庫，進

行選題工作，再編製成一份測驗，因此有越來越多學者投入選題策略問題的研究 [15,19,20,22,21,2]。

選題策略問題是屬於組合最佳化問題 (combinatorial optimization problem)，運算時間將隨著題庫題數增加呈指數倍成長 [17]，已有學者證明此為一 NP-hard 問題 [25]，而此類問題需要大量運算以求得最佳解，基因演算法 (Genetic Algorithm) 善於解決這類問題，此技術已被廣泛應用在科學與工程學問題上 [4,5,7,8,9,12,16,29]，而且也有學者將這種簡單基因演算法，應用於除了測驗題數與測驗訊息函數外，沒有額外限制式之選題策略問題上，得到不錯的效能 [2, 21]。然而，由於測驗需求的多樣性，選題常常牽涉到不同目的之限制式 [9,18,26,27,28]，這類問題也就是所謂的限制最佳化問題 (constrained optimization problem) [9]。在解決限制最佳化問題上，基因演算法中損失函數方法 (penalty function approach) 是最受歡迎的方式；它的優點是簡單、易於實施，而最大困難點在於能否找到適當的損失參數 (penalty parameter) [9,10]。因此，Kalyanmoy Deb 學者提出另一種演化技術來解決這個問題，它的特色在於：不需任何損失參數，並運用競爭式選擇運算子 (tournament selection operator) 成對比較，再利用歐幾里德距離和突變運算維持合理解多樣性，同時可達到更好的效能 [9]，但是，尚未有研究者將它運用於大量限制條件下之選題策略問題。

綜合上述，簡單基因演算法運用於選題策略上有不錯效能，然而在加入限制條件部分卻無深入且明確的探討。因此，本研究為了增加選題實用性，運用新進研究的演化技術於多限制條件之選題策略問題，並評估其效能及可行性。本文第二章，將介紹簡單基因演算法選題策略；第三章介紹 Deb 基因演算法選題策略；第四章則介紹進階基因演算法選題策略；第五章進行效能評估；最後，則做一結論與建議。

二、簡單基因演算法選題策略

簡單基因演算法應用於選題問題上時，將每個染色體字串 X^k 代表一份試卷 k ，含 n 個 bits(指題庫中的試題數 = n)，其中有 m 個要為 1(試卷的題數 = m)，其餘為 0。 x_i 代表試題 i 有沒有在這份試卷內 ($x_i = 1$ 表有； $x_i = 0$ 表無)。對每個染色體，我們可以計算目標測驗的訊息量，和新建立測驗的訊息量之間的誤差，這個**誤差平方和**，我們將它定義為演化過程適合度函數的一部分。

$$E(X^k) = \sum_{j=1}^s (d_j - o_j^k)^2 \quad (1)$$

其中

d_j 為能力等級 j 的目標訊息量

$j(=1 \sim s)$ 為能力等級； s 為能力等級數

$$o_j^k = \sum_{i=1}^n w_{ij} x_i^k, \forall j = 1, \dots, s$$

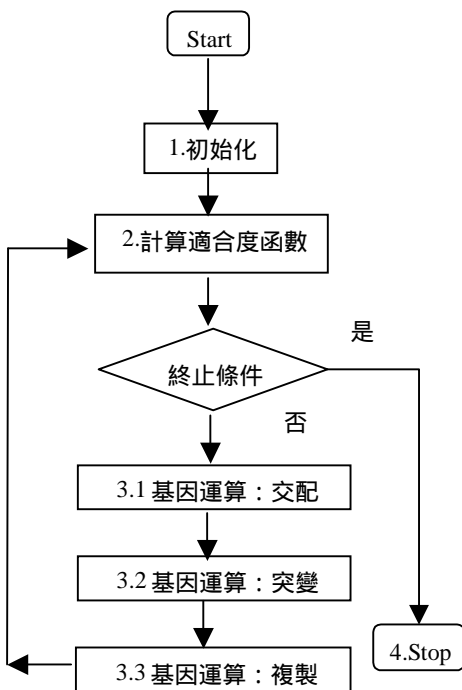
o_j^k 為染色體 k ，在能力等級 j 的測驗訊息量

$i(=1 \sim n)$ 為題庫第 i 題； n 為題庫總題數

w_{ij} 表示題目 i 在能力等級 j 的題目訊息量

$x_i^k \in \{0,1\}$ 在染色體 k 中，第 i 題選題狀態

藉著複製(reproduction)、交配(crossover)與突變(mutation)等基因運算，產生更佳的后代。演化過程如圖一：



圖一、基因演化流程

- (1) 設定染色體字串初始母群與演化過程的參數：

隨機產生由 n 個二元字串組成的染色體 P 個，每一染色體有 m 個為 1，其餘為 0。交配率、突變率和複製率分別以 p_c 、 p_m 、 p_r 表示，演化代數為 T ，初始值設為 0，最大值為 $gener_no$ 。

- (2) 計算母群內所有染色體的適合度函數，找出適合度最佳染色體 $chromosome_{opt}$ ：

$$\begin{aligned} fitness(X^k) &= E(X^k) + r^k \\ &= \sum_{j=1}^s (d_j - o_j^k)^2 + r^k \end{aligned} \quad (2)$$

r^k 為滿足 $\sum_{i \in C_q} x_i > k_q$ 或 $\sum_{i \in C_q} x_i < k'_q$ 的數量

其中

$E(X^k)$ 為目標函數

$q(=1 \sim p)$ 第 q 個限制式； p 為限制式總數

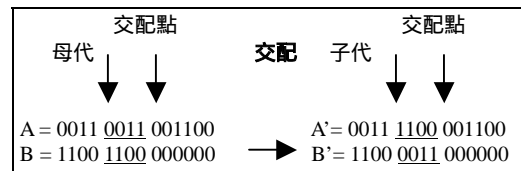
C_q 代表第 q 個限制式包含的題目集合

k_q, k'_q 第 q 個限制式的設定值

r^k 為違反限制式的數量

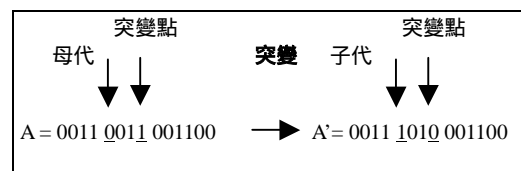
- (3) 對母群中每個染色體 k ，完成下列基因演算，產生 P 個子代。

(3.1) 採用兩點交配，每對母代隨著交配率 p_c 而產生子代，子代染色體字串有部分和母代相同，其餘則和另一母代相同。(如圖二)



圖二 交配過程

(3.2) 採用兩點突變(突變率為 p_m) 由母群隨機選取單一染色體，在單一染色體中再隨機選取兩個位置，0 變 1，1 變 0 的方式產生突變。(如圖三)



圖三 突變過程

(3.3) 採用的複製機制是將母群中的染色體經適合度計算排序後，保留適合度最佳的染色體，再依複製率 p_r 產

生數個相同的染色體，直接置入下一代的母群。

- (3.4) 如果子代中的染色體已符合測驗設計者的需求，或產生的子代數已達到最大值，即停止演化過程。否則“子代數”自動加 1，回到步驟(2)。

- (4) 停止演化。

最後，最佳的染色體字串就是選題問題的最佳解。

三、Deb 基因演算法選題策略

Kalyanmoy Deb 基因演算法是一個不需要損失參數的損失函數法，且可達到更佳效能之演化技術，但是尚未有人將此演化技術應用於大量限制條件之選題策略問題上。本方法主要是以簡單基因演算法為基礎，結合 Kalyanmoy Deb (Deb)的演化技術，運用於多限制條件的選題策略。與簡單基因演算法的主要差異有二：一為適合度函數的計算方式，另一為基因運算中的交配方式。交配方式中，當兩者皆為合理解時，會依設定的條件選取方式，進一步選取，而條件選取部分是以交配完兩個體間的距離及誤差大小為判別選取的標準。

進階基因演算法參數

Deb 基因演算法選題策略與進階基因演算法選題策略將運用進階基因演算法參數來改善選題的品質，參數的說明如下：

- (1) t 值：前置代數，即僅前 t 代做進一步的判斷選取。
 (2) Ed 值：交配完後兩合理解($P1, P2$)的歐幾里德距離。

$$Ed = \sqrt{\sum_{i=1}^n (x_i^{(P1)} - x_i^{(P2)})^2} \quad (3)$$

與簡單基因演算法選題策略主要相異之處，分述如下：

- (1) 適合度計算：

$$fitness(X^k) = \begin{cases} E(X^k), & \text{若合理解} \\ f_{\max} + r^k, & \text{否則} \end{cases} \quad (4)$$

$$= \begin{cases} \sum_{j=1}^s (d_j - o_j^k)^2, & \text{若合理解} \\ f_{\max} + r^k, & \text{否則} \end{cases}$$

r^k 為滿足 $\sum_{i \in C_q} x_i > k_q$ 或 $\sum_{i \in C_q} x_i < k_q'$ 的數量

f_{\max} 為母群中最差的合理解之目標函數值；
 若沒有合理解存在時， $f_{\max} = 0$

即當合理解時，適合度函數值等於目標函數值；非合理解時，適合度函數值則為母群中最差合理理解的目標函數值與違反限制式值的總和。

- (2) 基因運算 - 交配運算：

採用兩點交配的方式，再利用競爭式選擇由兩個不同的個體選擇其一，其比較法則如下：

- (2.1) 當兩個合理解做比較時，會依條件選取方式進一步選取。

if ($Ed < Ed_{\text{threshold}}$) *then* 選 $E(X^k)$ 小者
else
 前者與合理解集合一一求算 Ed
if ($Ed < Ed_{\text{threshold}}$) *then* 選後者, *else* 選前者

- (2.2) 當合理解與不合理解做比較時，會選擇合理解。

- (2.3) 當兩個不合理解比較時，會選擇限制式違反較少的不合理解。

四、進階基因演算法選題策略

進階基因演算法選題策略改良 Kalyanmoy Deb 基因演算法參數的使用，包含交配完新的兩個體間的距離 (Ed 值) 前置代數 (t 值) 等參數，因而提出了前置代數限制 (Restricted Generation: RG) 方法，即所謂的進階基因演算法選題策略。二者主要差別在於交配完皆為合理解時條件選取的部分，因此將只針對相異的部分做說明。

- (1) 基因運算 - 交配運算：

採用兩點交配的方式，再利用競爭式選擇由兩個不同的個體選擇其一，其比較法則如下：

- (1.1) 當兩個合理解做比較時，會依條件選取方式進一步選取。

if ($t < t_{\text{threshold}}$) *then*
if ($Ed < Ed_{\text{threshold}}$) *then* 選 $E(X^k)$ 小者
else
 前者與合理解集合一一求算 Ed
if ($Ed < Ed_{\text{threshold}}$) *then* 選後者, *else* 選前者
else 選 $E(X^k)$ 小者

- (1.2)當合理解與不合理解做比較時，會選擇合理解。
- (1.3)當兩個不合理解比較時，會選擇限制式違反較少的不合理解。

五、效能評估

由於線性規劃最常應用於找最佳化的問題，因此本研究將 SGA、RG、Deb 等演算法與新近研究 - 線性規劃(LP)做效能的比較，LP 的實驗結果主要參考自 [23]。為了使模擬的條件一致，使用之限制條件、基因演算法的基本設定等參數均與 LP 實驗模擬相同，說明如下。

題庫題數	1000 題 (如表一、表二)
測驗題數	30 題
演化代數	1000 代、1500 代
目標訊息量	單峰及雙峰各一組 (如表三、表四)
進階基因演算法參數	$Ed_{threshold} = 6$ $t_{threshold} = 75$
染色體	100 組
交配率	$p_c = 80\%$
突變率	$p_m = 10\%$
複製率	$p_r = 10\%$
限制條件	(內容說明如表五)

每個內容(1-10)領域題數 ≥ 2 , $\sum_{i \in c_q} x_i \geq 2$, $q = 1, \dots, 10, x_i \in \{0,1\}$

每個形式(1-5)領域題數 ≥ 4 , $\sum_{i \in c_q} x_i \geq 4$, $q = 11, \dots, 15, x_i \in \{0,1\}$

每個技巧(1-6)領域題數 ≥ 3 , $\sum_{i \in c_q} x_i \geq 3$, $q = 16, \dots, 21, x_i \in \{0,1\}$ (5)

q: 限制式(限制條件)編號

測驗長度的限制：題數 m 題, $\sum_{i=1}^n x_i = m$, $x_i \in \{0,1\}$,

n : 題庫題數, m : 試卷題數

題目長度的限制：每個題目長度 < 40 , $l_i x_i < 40$,

$i = 1 \dots n$, n : 題庫題數, l_i : 題目長度, $20 \leq l_i \leq 50$, $x_i \in \{0,1\}$

若以題庫中有 1000 題，要選取 40 題為例，限制條件共有： $10 + 5 + 6 + 1 + 1000 = 1022$ 個，所以，選出來之題目，必須滿足此 1022 個限制式。

今以各種基因演算法，各測試 10 次後，取其“誤差平方和”之平均值。

實驗結果：

1. SGA Deb RG 的誤差值都比 LP 來得小。
2. RG 的誤差值改善率佳，與 LP 相比平均改善率高達 99.96%。(如表六、表七)

六、結論與建議

本文中，我們改進簡單基因演算法中演化技術，改良參數使用策略，進行“多限制條件”選題策略問題的求解，並評估其效能及可行性。根據研究結果，做出如下的結論：

- (1)本文中運用的演算法，有效的解決“多限制條件”選題策略問題，求出的解有最小的誤差值，提供了更具實用性的技術，使測驗品質更能符合測驗設計者的需求。
- (2)以簡單的基因演算法為基礎，成功的結合 Kalyanmoy Deb 演化技術，並改良其參數使用策略，利用“Ed 值”增加多樣性及“t 參數”提高收斂至最佳值(或接近最佳值)的機會之特性，應用於演化過程中，成功的解決選題策略問題，大幅降低建構之測驗訊息函數與目標訊息函數之間的誤差值，並與其他新近研究 - LP 演算法比較，結果顯示本研究提出的前置代數限制基因演算法，相較於傳統常用的 LP 有更佳的结果，改善率高達 99.96%，與 Kalyanmoy Deb 所提之方法比較，改善率也高達 62.12%；不但擴展了基因演算法在多限制條件下的應用，也對教育評量者提供一個更有效的工具與技術。

誌謝

本研究計劃，經費由國科會科教處贊助，研究計劃編號：NSC 91-2520-S-024-014。

參考文獻

- [1]余民寧(民 81b)。題目反應的介紹 - 測驗理論的發展趨勢(二) 研習資訊, 9(1), 5-9。
- [2]孫光天、戴伯昌、賴膺守(民 88)。利用基因演算法於選題策略之研究。全國計算機會議(NCS' 99)論文集。私立淡江大學。
- [3]Ackerman, T. (1989). *An alternative methodology for creating parallel test forms using the IRT information function*. The Annual Meeting of the National Council for Measurement in Education, San Francisco.
- [4]Altenberg, L. (1994). The evolution of evolvability in genetic programming. In K. E. Kinnear, Jr., ed., *Advances in Genetic Programming*, London: MIT Press.

- [5]Axelrod, R. (1987). The evolution of strategies in the iterated Prisoner's Dilemma. In L. D. Davis, ed., *Genetic Algorithms and Simulated Annealing*. Morgan Kaufmann.
- [6]Baker, F. B., Cohen, A. S., & Barmish, B. R. (1988). Item characteristics of tests constructed by linear programming. *Applied Psychological Measurement*, 12(2), 189-199.
- [7]Bramlette, M. F. (1991). Initialization, mutation and selection methods in genetic algorithms for function optimization. In R. K. Belew and L. B. Booker, eds., *Proceedings of the Fourth International Conference on Genetic Algorithm*, Morgan Kaufmann.
- [8]Davis, L. (Ed.). (1991). *Handbook of Genetic Algorithms*. New York: van Nostrand Reinhold.
- [9]Deb, K. (2000). An efficient constraint handling method for genetic algorithms. *Computer Methods in Applied Mechanics and Engineering*, 186(2-4), 311-338.
- [10]Deb, K. (2001). *Genetic Algorithms for Optimization*. (Rep. No. 2001002). India: Department of Mechanical Engineering Indian Institute of Technology Kanpur, Kanpur Genetic Algorithms Laboratory(KanGAL).
- [11]De Gruijter, D. N. M. (1990). Test construction by means of linear programming. *Applied Psychological Measurement*, 14, 175-181.
- [12]Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley, Reading, Mass.
- [13]Hambleton, R. K., & Swaminathan, H. (1985). *Item Response Theory: Principles and Applications*. Hingham, MA: Kluwer, Nijhoff.
- [14]Lord, F. M. (1977). Practical applications of item characteristic curve theory. *Journal of Educational Measurement*, 14, 117-138.
- [15]Luecht, R. M., & Hirsch, T. M. (1992). Item selection using an average growth approximation of target information functions. *Applied Psychological Measurement*, 16(1), 41-51.
- [16]Mitchell, M. (1996). *An introduction to genetic algorithms*. London: MIT Press.
- [17]Papadimitrion, C. H., & Steiglitz, K. (1982). *Combinatorial Optimization: Algorithms and Complexity*. Prentice-Hall, Inc., NJ:Englewood Cliffs.
- [18]Sanders, P. F., & Verschoor, A. J. (1998). Parallel Test Construction Using Classical Item Parameters. *Applied Psychological Measurement*, 22(3), 212-223.
- [19]Stocking, M. L., & Swanson, L. (1993a). A method for severely constrained item selection in adaptive testing. *Applied Psychological Measurement*, 17(3), 277-292.
- [20]Stocking, M. L., & Swanson, L. (1993b). A model and heuristic for solving very large item selection problems, *Applied Psychological Measurement*, 17(2), 151-166.
- [21]Sun, K. T. (2000a). *A Genetic Approach to Parallel Test Construction*. International Conference on Computers in Education 2000, pp.83-90, The Grand Hotel, Taipei, Taiwan.
- [23]Sun, K. T. (2002). *A Genetic Algorithm for Parallel Test Forms* (Rep. No. 00101). Taiwan, Tainan: National Tainan Teachers College, AI & ICAI Lab.
- [24]Theunissen, T. J. J. M. (1985). Binary programming and test design. *Psychometrika*, 50, 411-420.
- [25]van der Linden, W. J. (1998). Optimal assembly of psychological and educational tests. *Applied Psychological Measurement*, 22(3), 195-209.
- [26]van der Linden, W. J., & Boekkooi-Timminga, E. (1989). A maximum model for test design with practical constraints. *Psychometrika*, 54, 237-247.
- [27]van der Linden, W. J., & Reese, L. M. (1998). A model for optimal constrained adaptive testing. *Applied Psychological Measurement*, 22(3), 259-270.
- [28]Wightman, L. F. (1998). Practical issues in computerized test assembly. *Applied Psychological Measurement*, 22(3), 292-302.
- [29]Wright, A. H. (1991). Genetic algorithms for real parameter optimization. In G. Rawlins, ed., *Foundations of Genetic Algorithms*. Morgan Kaufmann.
- [30]Wright, B. D., & Douglas, G. A. (1977). Best procedures for sample-free item analysis. *Applied Psychological Measurement*, 1, 281-295.
- [31]Wright, B. D., & Stone, M. H. (1979). *Best test design*. Chicago: MESA.

表一 電腦模擬 1000 題題庫(前 10 題)

題號	屬性	限制條件			題目參數			
		內容	形式	技巧	長度	a	b	c
1		8	3	4	28	1.464286	1.648441	0.102804
2		8	5	5	21	1.710872	2.175716	0.258096
3		4	5	6	21	2.889025	-0.815888	0.204974
4		8	1	4	34	1.455964	0.736180	0.229564
5		3	2	5	45	2.096159	2.916559	0.282193
6		3	4	6	27	1.974521	-2.361782	0.299883
7		7	1	4	23	1.026650	1.793306	0.156896
8		1	2	3	29	2.886856	2.878976	0.180275
9		3	1	1	40	1.702161	-0.523399	0.242546
10		4	4	2	25	2.083390	-2.515712	0.191594

代號意義：

一、 內容

- | | |
|-------|----------------|
| 1：天象 | 6：生物 |
| 2：光與熱 | 7：物質（固體、液體、氣體） |
| 3：電與磁 | 8：聲音 |
| 4：環境 | 9：化學 |
| 5：天氣 | 10：力 |

二、 形式

- | | |
|-------|-------|
| 1：選擇題 | 4：連連看 |
| 2：是非題 | 5：填充題 |
| 3：配合題 | |

三、 題目技巧

- | | |
|------|------|
| 1：知識 | 4：計算 |
| 2：記憶 | 5：理解 |
| 3：應用 | 6：推論 |

四、 題目參數

- a：鑑別度
b：難度
c：猜測度

表二 模擬題庫試題屬性

名稱	題目參數			限制條件			
	a	b	c	內容	形式	技巧	長度
範圍	0.8~3.0	-3.0~+3.0	0.1~0.3	1~10	1~5	1~6	20~50
類型	實數	實數	實數	整數	整數	整數	整數
平均值	1.916	-0.013	0.201	5.396	2.989	3.515	34.854
標準差	0.627	1.740	0.059	2.893	1.381	1.675	8.909

表三 測驗題數 30 題在單峰分布之目標訊息量屬性

	能力等級				
	1	2	3	4	5
能力值	-2.0	-1.0	0.0	1.0	2.0
訊息量	4	6	12	6	4

表四 測驗題數 30 題在雙峰分布之目標訊息量屬性

	能力等級				
	1	2	3	4	5
能力值	-2.0	-1.0	0.0	1.0	2.0
訊息量	4	10	4	10	4

表五 選題問題之限制條件

類別	
題目長度	Item length(number of words in the item)(Length)
測驗長度	Test length(total number of items in the test, Item Number)
測驗訊息函數	Test information function(TIF)
題目重疊數量	Number of overlapped items(Overlapped)
題目的曝光率	Limits on number of items used(Exposure Rate)
題目間之相依性	Mutually exclusive/dependent items(Dependence)
每個形式屬性的題數	Number of items per topic(Topic)
每個技巧屬性的題數	Number of items used skill(Skill)
每個內容領域的題數	Number of items per content category(Content)
每個認知層次的題數	Number of items per cognitive level(Level)

(引自 van der Linden & Reese, 1998)

表六 運用進階基因演算法之誤差值改善率(演化 1000 代)

項目 \ 演算法	RG	LP	SGA	Deb
單峰分布	0.000883	1.508440	0.021207	0.002153
雙峰分布	0.000873	2.551810	0.030765	0.003151
平均誤差	0.000878	2.030125	0.025986	0.002652
改善率 (%)	---	99.956754	96.621392	66.894796

改善率 (%) : $(error_x - error_{RG}) / error_x \times 100$
 $error_{RG}$: 運用 RG 方法產生的誤差值
 $error_x$: 運用 LP、SGA、Deb 方法產生的誤差值

表七 運用進階基因演算法之誤差值改善率(演化 1500 代)

項目 \ 演算法	RG	LP	SGA	Deb
單峰分布	0.000645	1.508440	0.009575	0.001338
雙峰分布	0.000770	2.551810	0.017542	0.002398
平均誤差	0.000708	2.030125	0.013558	0.001868
改善率 (%)	---	99.965147	94.781387	62.122591
改善率 (%) : $(error_x - error_{RG}) / error_x \times 100$ $error_{RG}$: 運用 RG 方法產生的誤差值 $error_x$: 運用 LP、SGA、Deb 方法產生的誤差值				