

改良式準隨機亂數株落演算法之組裝平行測驗

胡芸菁¹ 張庭毅² 呂易郡³ 鄭培成⁴

¹ 國立彰化師範大學 數位學習研究所

^{2,3} 國立彰化師範大學 工業教育與技術系

tychang@cc.ncue.edu.tw

⁴ 健行科技大學 資訊管理系

pccheng@uch.edu.tw

摘要

本研究採用 Chang 及 Shiu 提出的株落選擇演算法(CLONAL)進行改善,以株落選擇演算法為基礎,結合準隨機亂數(Quasi-Random Number Generator, QRNG)之平行測驗組裝方式,提出準隨機亂數株落演算法(QRNG CLONAL),除了可以同時建立多份平行試卷、解決循序組裝的不平等問題及減少限制式外,準隨機亂數機制可以進一步降低與目標測驗的均方誤差。經一系列實驗結果指出 QRNG CLONAL 不論在增加題庫試題數量、平行測驗數或有限制組裝時間的情況下,皆有利於組裝各種條件的平行測驗且有效維持測驗設計與品質。

關鍵詞: 株落選擇演算法、試題反應理論、準隨機亂數、選題策略、平行測驗。

Abstract

The simultaneously construct IRT-based parallel tests problem requires large numbers of variables and constraints, which leads to high computational complexities and now there is no polynomial time algorithm that exists for finding the optimal solution. Recently, Chang and Shiu proposed an adapted CLONAL algorithm to simultaneously construct IRT-based parallel tests. Their scheme is more efficient than its predecessors in terms of deviation and performance. In this paper, a Quasi-Random Number Generator is used for improving Chang and Shiu's scheme, called Q-CLONAL for short. The proposed Q-CLONAL not only avoids the inequality problem in the sequential construction, solves the drawback of larger numbers of variables, constraints in the simultaneous construction, but also has a better performance. Comparing with Chang and Shiu's CLONAL, under various situations (item bank size, the number of parallel tests, and the limit of execution time), the proposed Q-CLONAL is more suitable than Chang and Shiu's CLONAL in constructing IRT-based parallel tests..

Keywords: clonal selection algorithm, item response theory, parallel tests construction, quasi-random number generator, item selection.

1. 前言

測驗提供客觀的檢視、做出決定、評估受試者學習成果和能力,而測驗組合也因測驗目的與受試者的不同而有不同需求。近年來,電腦化適性測驗(Computerized Aptitude Tests, CAT)結合試題反應理論(Item Response Theory, IRT)被應用在各領域的大型測驗中,如托福、GRE 及 GMAT 等考試。在教育評量的相關研究中,「如何建構理想的測驗,以符合測驗設計者施測目的」,是過去許多學者探討的重要領域。隨著應用層面越來越廣,組裝高品質的 IRT 測驗變得越來越重要。

IRT 主要以個別試題的觀點,將受試者在某一試題上的表現情形與潛在特質透過一條連續性遞增的數學函數來表示,稱為試題特徵曲線(Item Characteristic Curve, ICC)。

組裝測驗(test construction)即是在某個教學及評量目標的引導之下,製作一組測量試題,以作為測量工具。組裝前,會先定義目標測驗,指在組裝測驗前訂定一份測驗規格作為組裝測驗目標,通常所考慮的準則包括內容、技巧、題型和作答時間,稱為多樣性限制條件,而組裝後所產生的測驗必須和目標測驗規格近乎相同。過去的研究中,組裝測驗的方法有:(1)隨機選取法(Random)、(2)標準法(Standard)、(3)中間難度法(Middle Difficulty)、(4)上下法(Up And Down)、(5)最大訊息法(Maximum Information)等[13],其執行時間短但其誤差大。從大量題庫中組出符合測驗需求的多樣性限制條件,為組合最佳化問題(combination optimization problem),屬於 NP-hard 問題。為了在有限時間內求得次佳解,因此學者利用概念啟發式演算法(Meta Heuristic Algorithm),例如類神經網路(Neural Network)[32]、蒙地卡羅演算法(Monte Carlo Algorithm)[4]、禁忌演算法(Tabu Search)[17]、免疫演算法(Immune Algorithm)[20]、粒子群最佳化演算法(Particle Swarm Optimization)[42]、蟻群演算法(Ant Algorithm)[21]、蜜蜂演算法(Bees Algorithm)[30]、基因演算法(Genetic Algorithm, GA) [11, 14, 16, 22, 25, 33]、株落選擇演算法(Clonal Selection Algorithm)[8]等方法分別應用在組裝測驗上,相較於傳統的組裝測驗方式,能在有限時間內,所組出之測驗與目標測驗有較小之誤差。

本研究利用準隨機亂數改善 Chang 及 Shiu[8] 所提出的株落選擇演算法(CLONAL)同時組裝平行測驗方式。CLONAL 是經由 GA 演變後的方法，因為 GA 本質上不適合用來同時組裝多份測驗，因此是以循序組裝的方式，而 CLOANL 不但可以同時組裝 5 份試卷，且能降低與目標 TIF 的誤差值，根據 Chang 及 Shiu 實驗指出株落選擇演算法相較於基因演算法平均誤差少 8.28 倍。本篇進一步在 CLOANL 的初始化階段加入蒙地卡羅法中的準隨機亂數，稱為 QRNG CLONAL，更能增加平行搜尋機制與獨立發展的特性，且減少同時組裝平行測驗限制式過多的問題。本論文架構如下，第二節介紹平行測驗的起源、特性和步驟。第三節介紹本研究採用的株落選擇演算法、多目標組裝測驗的問題模型和改良型準隨機亂數株落演算法的詳細步驟。第四節將介紹實驗方法與實驗結果。最後一節則針對本文做討論與分析。

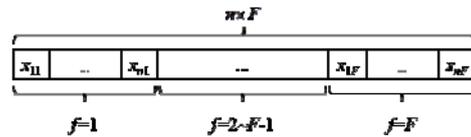
2. 相關研究

在實施測驗時，需要同時施測多份測驗的情況有：(1)週期性測驗、(2)給予不同受試者不同試卷以避免作弊、(3)包含前測和後測的二階段甚至多階段測驗、(4)多階段的適性測驗。目前大部分的平行測驗方法中，分為古典測驗理論(Classical Test Theory)與試題反應理論(Item Response Theory, IRT)。Armstrong 等人[2]使用古典測驗理論之試題指標，如鑑別度指標與難度指標，且結合 Network Flow 法再加上分數分佈指標(score distribution)、試題變異數(item variances)、試題與分數的相關(item-score correlations)等指標，提高了測驗平行的程度。由於 IRT 試題參數具有項目參數不變性、能力參數不變性與訊息函數功能，較優於古典試題理論，所以 IRT 試題參數逐漸成為主流。

平行測驗可分為循序組裝(sequential construction)與同時組裝(simultaneous construction)。Theunisse[36]使用數學規劃模型組裝 IRT 測驗時，採取隨機抽樣的組裝測驗方式，並且採用循序的方式進行，作法為組裝第一份測驗後，將題庫內用於組裝第一份測驗的試題排除，利用剩餘的試題組裝第二份測驗，用意是為了避免每份測驗試題重複。循序組裝測驗具備實作簡單與低運算量的優點，但組裝執行時間隨測驗多寡而增加，此外在組裝第一份試卷時，可能會將題庫內品質較好的試題先挑選出，導致優先組出的試卷品質最好，而後越來越差，這種情況稱為不平等問題(inequality problem)。若無法確保每張試卷品質一致，也無法測驗出受試者的學習情況，造成測驗間的偏差(bias)。

Armstrong 等人[2]的作法是在第一階段以循序的方式組裝多份平行測驗，第二階段將多份平行測驗的試題進行交換，以進一步降低整體誤差，但這樣的作法僅能稍微降低不平等現象所造成的誤

差。延伸兩階段組裝測驗的概念，Adema[1]提出一個兩階段同時組裝測驗的方法。其作法是在第一階段組裝一份題的試卷，為試卷試題數，為平行測驗數。第二階段再使用 LP 演算法將第一階段的題試卷分成 f 份相似的平行測驗，如圖一。其方法是用單一測驗模型去同時組裝平行測驗，因此不會有循序組裝的不平等問題，且不會有同時組裝變數與限制式過多的問題。雖然在第一階段能組裝出符合測驗規格的試卷，但在第二階段組裝測驗時不一定能確保平均測驗品質，因此仍有很大的改善空間。



圖一. Adema 的測驗編碼方式

過去只有少數研究能不需加入大量的變數和限制式且能解決不平等問題，但這些方法組裝的測驗不足以符合目標測驗規格或計算量太大，無法在短時間內同時產生多份低誤差的平行測驗[8]。由於測驗組裝是一種組合最佳化問題，試題組合會隨著試題數量增加而呈指數成長[38]。所以近年來朝向採用啟發式演算法，先前提到的啟發式演算法都是針對單一測驗組裝，也應用於平行測驗，像是模擬降溫法、Weighted Deviations model and Heuristic (WDH)法、基因演算法、株落演算法等。Stocking 等人[31]使用 WDH 法，作法分為兩階段，第一階段為組裝兩份測驗 a 和 b，針對題庫內每一題計算試題訊息量，選擇與題庫所有題目的平均試題訊息量距離最小者，交叉分配給測驗 a 和測驗 b；第二階段使兩份測驗的試題盡量相似，隨機從測驗 a 中選出一題，和測驗 b 中的所有試題訊息量差距最大者設為 x ，再從剩餘試題中選出一題，和測驗 b 中所有試題訊息量差距最大者設為 y ，如果 $y < x$ ，則交換兩題。相反地，將上述步驟測驗 b 取代為測驗 a，重複步驟直到沒有交換為止。此方法雖可有效的找出最好的組合，但其缺點是反覆執行，花費時間長。Wang 等人[41]將 Stocking 等人[31]的方式作進一步的改良，根據與目標訊息量的差距來選擇題目，選擇試卷中與備選試題差異最大的作交換。以上兩種方式皆基於遵循尋找目標和交換模式的啟發式算法，雖然可以有效的得到平行測驗，但可能造成無法收斂或無法找出最接近目標測驗的結果。Sun 等人[33]提出使用二進制基因演算法，世代數為 2000 的最佳情況下，均方差為 0.5321。在相同測驗規格下，和 Stocking 等人[31]與 Wang 等人[41]相比，改善率平均提升 95%。Hwang 等人[15]則提出實數型編碼的禁忌演算法來進行同時組裝平行測驗，其作法是在組裝時將多份測驗同時組裝在一起，再分成多份試卷，這樣可以避免試題重複的情形發生，也可以減少限制式，由於使用的是實數型，編碼長度較二進制短，因此可以降低不少執行時間。總結來說，同時測驗組裝雖然能有效的克服循序法的不平等問題[1, 6, 7, 15, 23]，但提高了問

題的複雜度導致增加平行測驗間的誤差。亦有學者朝向硬體發展，藉由電腦平行運算的技術組裝平行測驗，Songmuang 及 Ueno[30]應用蜜蜂演算法，使用多個處理器和多台電腦同時組裝測驗，以減少計算時間，但因設備成本太高，較不適合廣泛使用於組裝測驗系統上。

以往的組裝測驗研究幾乎都著重於相同目標測驗訊息量(Target Information Function, TIF)的平行測驗，根據 McDonald 的定義稱為 TSIF-Parallel tests，又稱弱平行測驗(weakly parallel tests)，有相同測驗特性的平行測驗則稱為 TCC-Parallel tests。當平行測驗同時符合 TSIF-Parallel tests 與 TCC-Parallel tests 就可提升平行測驗間相互比較的精確度。然而，除了 TIF 外，測驗特性也需完全相同，如測驗試題數、內容、題型和長度。由於符合兩種 TCC 和 TSIF 的平行測驗難度很高，Chang 及 Shiu[8]提出的株落選擇演算法於組裝平行測驗，藉由 CLONAL 的平行搜尋機制來解決限制式過多的問題，和 Sun[33]的 GA 相較之下，CLONAL 較 GA 的均方差平均改善了 8.71 倍。

Sun 等人[33]採用的 GA 和 Chang 及 Shiu[8]採用的 CLONAL 在組裝平行測驗中，其初始化步驟皆是使用一般虛擬亂數產生器(Pseudo Random Number Generator, PRNG)[29]作為試題編碼，PRNG 所產生的亂數會有叢聚的問題，產生隨機樣本不滿足均勻分布，造成題庫中的某些題目不會被選中的情形。本研究將採準隨機亂數(Quasi-Random Number Generator, QRNG)，如圖三，可看出 QRNG 比 PRNG 分布較均勻。QRNG 又可稱為低差異數列(Low-Discrepancy Sequence, LDS)，是一種超均勻分布的隨機亂數，此種方法是給定一個規則產生數列，轉換後所得到的值大致上會落在(0, 1)的範圍內並且均勻分佈，所以可以快速的達到收斂，並可以節省大量的計算時間。近來有許多學者提出改進的方法[10, 12, 27, 35]，使得執行效率和準確率遠高於原始隨機亂數法。現代普遍常用的分為兩種，Halton sequences 和 Sobol sequences。

Maaranen 等人[24]在基因演算法的初始化步驟使用 QRNG，證實可以提升實驗結果的準確性，且可減少世代數。Reese[26]和 Kimura, Matsumura[19]進一步發現 QRNG 應用在 GA 解決非線性或多目標問題時可以提高搜尋效能。因此，本篇將 CLONAL 結合 QRNG 方法，CLONAL 可將分散在搜尋空間中的所有解，各自找出了所在區域中的局部最佳解[8]，再加上許多學者證實 QRNG 可提高搜尋效能，降低誤差，因此能有效提升組裝平行測驗的效果。

3. 株落選擇演算法

株落選擇演算法是 De Castro 及 Von Zuben[9]提出的啟發式演算法，此演算法是模擬生物免疫系統中的內容，並以生物學的觀點來模擬免疫行為，稱為人工免疫系統。此演算法是以其中的株落選擇概念衍生而成，有別於其他演算法，它利用族群相

似程度的關係，增加族群的多樣性，以避免陷入局部最佳解的可能性，使得在求解空間的蒐尋過程中，能夠快速收斂且找到全域最佳解。

B 淋巴細胞是用來生產抗體(antibody)，當病毒、細菌、不同種的血液、移植的器官等出現時就是所謂的抗原(antigen)，B 淋巴細胞的抗體與抗原間的反應強度衡量稱為親和度(affinity)，親和度越高表示抗體越能有效對抗抗原，然後挑選出高親和力的抗體，B 淋巴細胞會迅速製造(增殖)一個或多個抗體，以擊退抗原，這些被增殖出的細胞就稱作克隆(clone)，再進行突變(mutation)，突變有利於增加親和度，此過程為株落選擇(clonal selection)，而這些高親和度的抗體將會越來越多，與抗原的親和力也會越來越高。

應用於測驗組裝的問題，每個抗體代表的是一份測驗，透過株落選擇的方式(增殖、突變)以滿足目標測驗的規格，目標測驗指的就是抗原，不斷重覆以上過程，直到產生最符合目標測驗規格(即親和力最高)的試卷為止。

主要步驟：

- (1)產生初始族群(initialization)：以隨機的方式產生初始抗體族群($popu_size$)，每一個抗體即代表一份測驗。
- (2)計算親和力(calculate affinity)：計算抗體與抗原間的適合程度。在測驗組裝問題中，親和力代表組裝測驗與目標測驗之間的差距，差距越小表示親和力越高。
- (3)增殖(proliferation)：選出較高親和力 n_{best} 個抗體來進行增殖，增殖出的抗體稱為克隆群體 C_{group} ，而克隆群體內的抗體數量必須和原抗體的親和力成正比。
- (4)突變(mutation)：將克隆群體進行突變，被突變的抗體由突變率決定，而突變率和親和力成反比，換句話說，親和力越低，突變率越高。
- (5)重新選擇(re-selection)：重新計算克隆群體中所有抗體的親和力，然後選出 n_{best} 個親和力最高的抗體回族群取代原抗體。
- (6)重組(meta-dynamic)：從族群中選出 $popu_size - n_{best}$ 個親和力最低的抗體，隨機產生新的抗體做取代。
- (7)循環(cycle)：反覆地執行步驟 2 到 6，直到完成指定的世代數。

株落演算法的搜尋機制是採用平行搜尋，族群中的每個解都是獨立發展，不會朝族群中最佳解的方向移動。到了搜尋後期所有解會分散在搜尋空間中，並各自找出所在區域中的局部最佳解。由於具備了獨立發展的特性且保有族群的多樣性，因此適合用於必須同時找出多個解的問題，非常符合平行測驗組裝的需求。

3.1 多目標組裝測驗的問題模型

本研究定義一個多目標測驗問題(Test Sheet

Composition Problem, TSC), 令題庫中有 N 個試題, 分別為 x_1, x_2, \dots, x_N , 每題皆對應一種章節、技巧、題型, 則可產生一份包含 C 個章節、 S 種技巧和 Y 種題型的試卷, 而 F 份試卷裡的所有題目皆不重覆。多目標平行組裝測驗的目的是建構出多份條件組合皆相同的試卷。以實數型表示模型問題, 其定義如下:

$$\text{Minimize } \sum_{k=1}^K \sum_{i=1}^n I_{x_i}(\theta_k) \quad (1)$$

Subject to

$$\sum_{i=1}^n I_{x_i}(\theta_k) \geq D(\theta_k), \quad k=1, \dots, K \quad (2)$$

$$\sum_{i=1}^n c_{x_i p} = C_p, \quad p=1, \dots, P \quad (3)$$

$$\sum_{i=1}^n s_{x_i q} = S_q, \quad q=1, \dots, Q \quad (4)$$

$$\sum_{i=1}^n y_{x_i r} = Y_r, \quad r=1, \dots, R \quad (5)$$

$$\sum_{i=1}^n u_{x_i} \leq U \quad (6)$$

$$x_i \in [1, N], \quad i=1, \dots, N \quad (7)$$

x_i	實數型變數, 表示題庫裡的試題編號
n	目標測驗長度
$I_{x_i}(\theta_k)$	試題 x_i 在能力等級 k 時的試題訊息量
$D(\theta_k)$	等級 k 時的目標測驗訊息量
$c_{x_i p}$	二進制變數, 若為 1, 表示試題 x_i 屬於內容 p , 反之則否
C_p	章節 p 的目標測驗題數
$s_{x_i q}$	二進制變數, 若為 1, 表示試題 x_i 屬於技巧 q , 反之則否
S_q	解題技巧屬性 q 的目標測驗題數
$y_{x_i r}$	二進制變數, 若為 1, 表示試題 x_i 屬於題型 r , 反之則否
Y_r	題型屬性 r 的目標測驗題數
u_{x_i}	表示試題 x_i 的字數
U	表示目標測驗字數

目標式(1)和限制式(2)表示在每個能力等級 k 試題訊息量的總和必須大於等於目標訊息量; 限制式(3)、(4)及(5)分別代表試題在各章節 p 、解題技巧 q 及題型 r 的題數必須符合目標題數; 限制式(6)表示測驗字數總和必須小於目標字數上限; 限制式(7)表示 x_i 為任意實數, 即實數型編碼。下一節詳細介紹如何使用 QRNG 株落演算法來解決此組裝測驗模型。

3.2. QRNG 株落演算法

本文根據 Chang 及 Shiu[8]提出的 CLONAL, 其初始族群步驟所使用的虛擬亂數產生器(PRNG)

以準隨機亂數(QRNG)取而代之, PRNG 普遍用於一般電腦程式, 其原理是利用序列或等式, 產生 0 到 1 之間的資料組, 但 PRNG 無法完全符合隨機變數的統計特性。針對上述問題, 許多研究提出使用 quasi-random 改善均勻度不佳, 與計算結果不準確和收斂速度慢的問題。因此, 我們提出準隨機亂數(QRNG)改良 CLONAL, 由於 quasi-random 產生的隨機數為均勻分佈, 可以解決分佈不均的問題。

由於在產生初始族群、突變、重組三個階段有可能會導致試題重疊, 解決方法是在演算法中加入排除機制, 當試題被挑選到測驗時, 則此試題會暫時列入排除試題區 *ban_list*, 演算法挑選剩餘的試題, 這個做法可以避免試題重覆, 並且可同時組出多份測驗。以下將介紹 QRNG 株落演算法。

主要步驟:

(1) 產生初始族群: 使用 QRNG[28] 隨機產生 *popu_size* 個抗體(測驗), 每個抗體包括 n 個實數 x_i , 而 $x_i \in [1, N]$, 被選擇的試題暫時移到 *ban_list* 中, 以避免試題被重覆挑選。

QRNG:

首先, 選取 d 個隨機方向數, 再以不同的本質多項式遞迴產生其他的方向數, 其中 d 代表本質多項式的階數, 利用方向數產生 QRNG 亂數, 以下詳述 QRNG 產生過程:

(1.1) 計算本質多項式:

令任意十進制奇數 B , 將 B 轉換為二進制 $a_0, \dots, a_d \in \{0, 1\}$, 形成本質多項式 P :

$$P = a_0 x^d + a_1 x^{d-1} + \dots + a_{d-1} x^1 + a_d x^0 \quad (8)$$

(1.2) 求方向數:

方向數由等式(9)遞迴計算產生, 其中 \oplus 為 XOR 運算。為了方便計算, 假設 $v_i = m_i / 2^i$ 簡化為等式(10)。當 $i \leq d$ 時無法由等式產生 m_i , 當 $i > d$ 時需假設 m_1, \dots, m_d 為任意奇數並滿足 $0 < m_i < 2^i$ 。將 m_1, \dots, m_d 代入等式(15), 由 $v_i = m_i / 2^i$ 推得方向數 v_i 。

$$(v_i)_2 = (a_1 v_{i-1} \oplus a_2 v_{i-2} \oplus \dots \oplus a_{d-1} v_{i-d+1} \oplus v_{i-d} \oplus [v_{i-d} / 2^d])_2, \quad i > d \quad (9)$$

$$m_i = \begin{cases} 2k-1, k \in N, \text{ for } \forall i \leq d, 0 < m_i < 2^i \\ 2a_i m_{i-1} \oplus 2^2 a_{i-1} m_{i-2} \oplus \dots \oplus 2^{d-1} a_{d-i+1} m_{i-d+1} \oplus 2^d m_{i-d} \oplus m_{i-d}, \text{ for } \forall i > d \end{cases} \quad (10)$$

(1.3) 遞迴產生 QRNG 亂數:

根據等式(11)遞迴產生 $x(n)$ 數列, 其中 $0 \leq n \leq 2^N - 1$, N 為任意正整數, n 以二進制表示時, 由右至左搜尋第一個為 0 的位數為 k , 由等式(14)得到 v 等式(11)遞迴產生 QRNG 亂數。

$$x(n) = \begin{cases} N, \text{ for } n = 0 \\ x(n-1) \oplus v_k, \text{ for } 1 \leq n \leq 2^N - 1 \end{cases} \quad (11)$$

(2) 計算親和力: 親和力是用來辨識抗體與抗原間的吻合度。本研究在親和力的計算上是使用測驗與目標測驗的誤差值為評估方式, 等式(12)到等式(16), 根據問題模型來計算每份測驗的章節、技巧、題型、字數與目標測驗規格的誤差, 等式(17)將所有

誤差乘上權重後全部相加即為親和力，權重為自訂的值，可依自認為重要性而作調整，而偏差值 E_I 、 E_C 、 E_Q 、 E_R 、 E_U 為 0 代表測驗完全符合目標測驗規格，表示親和力越高，越接近目標測驗。

$$\sum_{k=1}^K \left| \sum_{i=1}^n I_{x_i}(\theta_k) - D(\theta_k) \right| = E_I \quad (12)$$

$$\sum_{p=1}^P \left| \sum_{i=1}^n c_{x_i p} - C_p \right| = E_C \quad (13)$$

$$\sum_{q=1}^Q \left| \sum_{i=1}^n s_{x_i q} - S_q \right| = E_Q \quad (14)$$

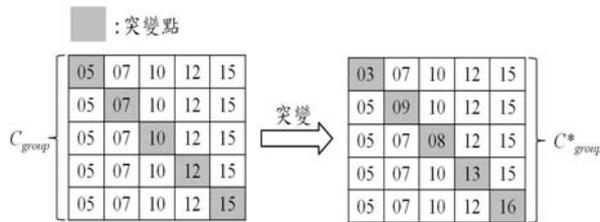
$$\sum_{r=1}^R \left| \sum_{i=1}^n y_{x_i r} - Y_r \right| = E_R \quad (15)$$

$$\max(0, \sum_{i=1}^n u_{x_i} - U) = E_U \quad (16)$$

$$\text{affinity} = W_I E_I + W_C E_C + W_Q E_Q + W_R E_R + W_U E_U \quad (17)$$

(3) 增殖：選擇親和力較高的 n_{best} 個抗體進行增殖，因為本文突變設計的關係，所以克隆數量設為 n ，因此每個克隆群體 C_{group} 中有 n 個抗體，總共有 n_{best} 個抗體。

(4) 單點突變：以 C_{group} 為單位進行單點突變，以圖二為例，假設一份測驗有 5 題試題，分別為 x_1, x_2, x_3, x_4, x_5 (題號 5、題號 7、題號 10、題號 12、題號 15)，因此克隆族群 C_{group} 為 5 個複製的抗體，第一個抗體突變 x_1 (題號 5)， x_1 從已經排除原本族群內試題的題庫中隨機選出一題(題號 3)作取代，突變後也將題號 3 這題納入 ban_list ，第二個抗體突變 x_2 (題號 7)，依此類推，突變後的 C_{group} 稱為 C_{group}^* 。



圖二. 單點突變範例

(5) 重新選擇：計算 C_{group}^* 內 n 個抗體和 C_{group} 原始抗體的親和力互相做比較，保留最佳親和力的抗體取代原始抗體，如果 C_{group}^* 內的抗體親和力比 C_{group} 原始抗體差，則保留原始抗體。

(6) 重組：從族群中挑選出親和力最差的 $popu_size - n_{best}$ 個抗體並用全新的隨機抗體取代，在產生隨機抗體時，同時也排除整個族群的試題再做隨機選擇。

(7) 循環：重覆步驟(2)到步驟(6)，直到完成指定的世代數 gen 。而其中挑選出前 F 個親和力較佳的抗體，即為 F 份平行測驗。

4. 實驗設計與結果

由於 Chang 及 Shiu[8]所提出的 CLONAL 較過去所有研究[33, 34, 41]有較佳的結果，因此本節針

對和 Chang 及 Shiu 的 CLONAL 進行一系列實驗比較。本篇以 Chang 及 Shiu 的實驗設計題庫，題庫試題數量 16000 題，作為組裝測驗依據。題目參數鑑別度介於 0.8~3.0，困難度介於 -3.0~3.0，猜測度介於 0.1~0.3，章節分為 1~10 章，技巧分為 1~6 種(計算、知識、記憶、推理、理解、應用)，題型分為 1~5 種(是非題、單選題、多選題、填充題、簡答題)，字數介於 20~50 字。

為了評估 QRNG CLONAL 是否能滿足測驗規格並且和 CLONAL 各作一百次的組裝，每次組裝的測驗目標都是隨機產生。本文考慮兩種組卷目標 TIF 波型，分別是單峰型和雙峰型。對單峰型 TIF，目標訊息量的隨機產生範圍為 4~5, 6~8, 18~21, 6~8, 4~5，分別對應到能力等級。而對雙峰型 TIF，隨機產生範圍為 7~8, 13~15, 9~11, 13~15, 7~8，分別對應到能力等級。章節、技巧、題型、測驗字數的目標題數也是隨機產生，而訊息量偏差值的精確率設為小數點後四位數，因此設 $W_I=W_C=W_Q=W_R=W_U=10000$ 。

在題庫為 16000 情況下，同時組裝 20 份評測測驗，CLONAL 和 QRNG CLONAL 執行一百次的平均實驗結果，誤差為和目標測驗間的差距，均方差代表的是平行測驗間的品質差距。表一顯示 QRNG CLONAL 改善率可達 45.3%與 33.6%。

表一. CLONAL 和 QRNG CLONAL 實驗結果(20 份)

題數	CLONAL		QRNG CLONAL	
	16000			
TIF*	S	D	S	D
誤差	1.30	1.24	0.98	0.97
均方差	0.76	0.65	0.41	0.43
時間(分)	7.56	7.66	7.80	7.88
改善率**	-	-	45.3	33.6

*S=單峰 TIF, D=雙峰 TIF **改善率=(QRNG CLONAL 均方差-CLONAL 均方差)/CLONAL 均方差*100 (%)

5. 結論

先前的研究證明了啟發式演算法適合解決平行測驗問題[40]，本文提出了 QRNG 改良株落演算法，採用準隨機亂數將株落選擇演算法提升了平行測驗間的品質差異。並且從龐大的題庫中，將平行組裝測驗問題簡單化，解決了大量的選擇變數和多樣性限制條件所造成測驗品質下降的問題。實驗證明了和 Chang 及 Shiu 提出的 CLONAL 相較之下，QRNG CLONAL 的平行測驗的品質更好，且降低多份測驗間的差異，和目標規格也有更小的偏差。整體來說，使用 QRNG CLONAL 組裝測驗能解決循序組裝的不平等問題，有效降低問題複雜度，能在可接受的時間內組裝符合目標規格的測驗，當擴充題庫或平行試卷數，能保持測驗間的差異穩定度。

誌謝：非常感謝國科會計畫編號：

NSC101-2622-E-018-001-CC3

參考文獻

- [1] J. J. Adema, "Implementations of the branch-and-bound method for test construction problems," *Methodika*, 1992.
- [2] R. D. Armstrong, D. H. Jones, and Z. Wang, "Automated parallel test construction using classical test theory," *Journal of educational and behavioral statistics*, vol. 19, no. 1, pp. 73-90, 1994.
- [3] F. B. Baker, A. S. Cohen, and B. R. Barmish, "Item characteristics of tests constructed by linear programming," *Applied Psychological Measurement*, vol. 12, no. 2, pp. 189-199, 1988.
- [4] D. I. Belov, and R. D. Armstrong, "Monte Carlo test assembly for item pool analysis and extension," *Applied Psychological Measurement*, vol. 29, no. 4, pp. 239-261, 2005.
- [5] E. Boekkooi-Timminga, "The construction of parallel tests from IRT-based item banks," *Journal of educational and behavioral statistics*, vol. 15, no. 2, pp. 129-145, 1990.
- [6] E. Boekkooi-Timminga, "Simultaneous test construction by zero-one programming," Department of Education of the University of Twente, 1986.
- [7] K. Breithaupt, A. Ariel, and B. P. Veldkamp, "Automated simultaneous assembly for multistage testing," *International Journal of Testing*, vol. 5, no. 3, pp. 319-330, 2005.
- [8] T. Y. Chang, and Y. F. Shiu, "Simultaneously construct IRT-based parallel tests based on an adapted CLONALG algorithm," *Applied Intelligence*, vol. 36, no. 4, pp. 979-994, 2012.
- [9] L. N. De Castro, and F. J. Von Zuben, "Learning and optimization using the clonal selection principle," *Evolutionary Computation*, *IEEE Transactions on*, vol. 6, no. 3, pp. 239-251, 2002.
- [10] H. Faure, "Discrèpance de suites associées a un système de numération (en dimension s)," *Acta Arith*, vol. 41, no. 4, pp. 337-351, 1982.
- [11] M. Finkelman, W. Kim, and L. A. Roussos, "Automated test assembly for cognitive diagnosis models using a genetic algorithm," *Journal of educational measurement*, vol. 46, no. 3, pp. 273-292, 2009.
- [12] J. H. Halton, "On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals," *Numerische Mathematik*, vol. 2, no. 1, pp. 84-90, 1960.
- [13] R. K. Hambleton, and H. Swaminathan, "Item response theory: Principles and applications," Kluwer-Nijhoff Pub.(Boston and Hingham, MA, USA), 1985.
- [14] J. H. Holland, "Adaptation in natural and artificial systems, University of Michigan press," Ann Arbor, MI, vol. 1, no. 97, pp. 5, 1975.
- [15] G.-J. Hwang, H.-C. Chu, P.-Y. Yin, and J.-Y. Lin, "An innovative parallel test sheet composition approach to meet multiple assessment criteria for national tests," *Computers & Education*, vol. 51, no. 3, pp. 1058-1072, 2008.
- [16] G. J. Hwang, B. M. T. Lin, H. H. Tseng, and T. L. Lin, "On the development of a computer-assisted testing system with genetic test sheet-generating approach," *Systems, Man, and Cybernetics, Part C: Applications and Reviews*, *IEEE Transactions on*, vol. 35, no. 4, pp. 590-594, 2005.
- [17] G. J. Hwang, P. Y. Yin, and S. H. Yeh, "A tabu search approach to generating test sheets for multiple assessment criteria," *Education*, *IEEE Transactions on*, vol. 49, no. 1, pp. 88-97, 2006.
- [18] H. Jeng, and S. Shih, "A comparison of pairwise and group selections of items using simulated annealing in automated construction of parallel tests," *Psychological Testing*, vol. 44, no. 2, pp. 195-210, 1997.
- [19] S. Kimura, and K. Matsumura, "Genetic algorithms using low-discrepancy sequences," *Proceedings of the 2005 conference on Genetic and evolutionary computation*, pp. 1341-1346, 2005.
- [20] C. L. Lee, C. H. Huang, and C. J. Lin, "Test-sheet composition using immune algorithm for E-learning application," *New Trends in Applied Artificial Intelligence*, pp. 823-833, 2007.
- [21] W. L. Leite, I.-C. Huang, and G. A. Marcoulides, "Item selection for the development of short forms of scales using an ant colony optimization algorithm," *Multivariate Behavioral Research*, vol. 43, no. 3, pp. 411-431, 2008.
- [22] F. T. Lin, C. Y. Kao, and C. C. Hsu, "Applying the genetic approach to simulated annealing in solving some NP-hard problems," *Systems, Man and Cybernetics*, *IEEE Transactions on*, vol. 23, no. 6, pp. 1752-1767, 1993.
- [23] W. J. Linden, and J. J. Adema, "Simultaneous assembly of multiple test forms," *Journal of educational measurement*, vol. 35, no. 3, pp. 185-198, 1998.
- [24] H. Maaranen, K. Miettinen, and M. M. Mäkelä, "Quasi-random initial population for genetic algorithms," *Computers & Mathematics with Applications*, vol. 47, no. 12, pp. 1885-1895, 2004.
- [25] M. Melanie, "An introduction to genetic algorithms," Cambridge, Massachusetts London, England, Fifth printing, vol. 3, 1999.
- [26] A. Reese, "Random number generators in genetic algorithms for unconstrained and constrained optimization," *Nonlinear Analysis: Theory, Methods & Applications*, vol. 71, no. 12, pp. 679-692, 2009.
- [27] I. Sobol, "Multidimensional quadrature formulas and Haar functions," *Izdat" Nauka"*, Moscow, 1969.
- [28] I. Sobol, "Quasi-monte carlo methods," *Progress in Nuclear Energy*, vol. 24, no. 1, pp. 55-61, 1990.
- [29] I. Sobol, and B. Shukhman, "Quasi-random points keep their distance," *Mathematics and Computers in Simulation*, vol. 75, no. 3, pp. 80-86, 2007.
- [30] P. Songmuang, and M. Ueno, "Bees algorithm for construction of multiple test forms in e-testing," *Learning Technologies*, *IEEE Transactions on*, vol. 4, no. 3, pp. 209-221, 2011.
- [31] M. L. Stocking, L. Swanson, and M. Pearlman, "Application of an automated item selection method to real data," *Applied Psychological Measurement*, vol. 17, no. 2, pp. 167-176, 1993.
- [32] K. Sun, and S. Chen, "A study of applying the artificial intelligent technique to select test items," *Psychological Testing*, vol. 46, no. 1, pp. 75-88, 1999.
- [33] K. T. Sun, Y. J. Chen, S. Y. Tsai, and C. F. Cheng, "Creating IRT-based parallel test forms using the genetic algorithm method," *Applied measurement in education*, vol. 21, no. 2, pp. 141-161, 2008.
- [34] L. Swanson, and M. L. Stocking, "A model and heuristic for solving very large item selection problems," *Applied Psychological Measurement*, vol. 17, no. 2, pp. 151-166, 1993.
- [35] S. Tezuka, "Polynomial arithmetic analogue of Halton sequences," *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, vol. 3, no. 2, pp. 99-107, 1993.
- [36] T. Theunissen, "Binary programming and test design," *Psychometrika*, vol. 50, no. 4, pp. 411-420, 1985.
- [37] W. J. Van der Linden, "Linear models for optimal test design," Springer Science+ Business Media, 2005.
- [38] W. J. Van der Linden, "Optimal assembly of psychological and educational tests," *Applied Psychological Measurement*, vol. 22, no. 3, pp. 195-211, 1998.
- [39] W. J. van ver Linden, and E. Boekkooi-Timminga, "A maximin model for IRT-based test design with practical constraints," *Psychometrika*, vol. 54, no. 2, pp. 237-247, 1989.
- [40] A. J. Verschoor, "Genetic algorithms for automated test assembly," University of Twente, 2007.
- [41] C. Wang, and T. Ackerman, "Two item selection algorithms for creating weakly parallel test forms using the IRT information functions," *Psychological Testing*, vol. 44, no. 2, pp. 123-140, 1997.
- [42] P. Y. Yin, K. C. Chang, G. J. Hwang, G. H. Hwang, and Y. Chan, "A particle swarm optimization approach to composing serial test sheets for multiple assessment criteria," *journal of educational technology and society*, vol. 9, no. 3, pp. 3, 2006.