

使用貪婪演算法作為一有效益之選題策略

孫光天

陳岳宏

賴膺守

謝凱隆

陳新豐*

ktsun@ipx.ntntc.edu.tw

國立台南師範學院 資訊教育研究所

*國立政治大學教育學研究所

摘要

在教育評量技術中，題目反應理論(Item Response Theory: IRT)，克服了許多在傳統教育與心理測驗中對測驗建構、評量與使用的問題，為數不少的研究與高品質測驗之建構均以此理論完成；然而，一般選題策略的方法，雖然很有效率卻缺乏彈性，以至於建構之測驗很難趨近設計者之需求。本文中，我們將提出一貪婪演算法做為選題策略，其時間(運算)複雜度與傳統方法相同，但選題結果之測驗訊息量與設計者期望訊息量之間誤差，遠低於早期傳統方法，並優於其他新近研究結果，肯定了本研究所提之技術，對測驗發展技術給予一新的研究方向。

關鍵字：題目反應理論，選題策略，貪婪演算法

1. 簡介

七〇年代起，題目反應理論[13,14]成為測驗學者們研究的焦點，而這新的測驗理論被認為可以解決古典測驗理論所無法解決的幾個問題，並且由於現代資訊科技的迅速發展，使適性測驗可具實用價值，因此，結合題目反應理論、適性測驗、電腦科技等理論與技術的電腦化適性測驗，在國內正在被熱烈的討論[1,2,4,5,6,7]。然而，一般紙筆測驗仍被大量採用，因此，如何由題庫中選取題目，以符合測驗設計者欲達成之目標，將是一項重要且有意義的研究，此外，欲建立多份平行測驗，以評估教學活動之效能時，選題策略更形重要。因此，設計一個高效益的選題策略，以符合測驗設計者施測需求，為本研究之重點。題目反應理論其特點在於(1)題目參數估計與受試者彼此互相獨立(independent)，(2)受試者能力估計與施測題目彼此互相獨立，及(3)能力估計的精確度是可估計(已知)的；因此，廣泛應用於教育評量之相關研究中，其中一重要研究領域為：「設計一個高效益的選

題策略，以符合測驗設計者施測需求」，亦為多人所探討；例如：要建構一份學術獎勵之成就測驗，則須由題庫中選取較難的題目來施測，才能精準選出高能力學生，因此，測驗試題之選取，為一重要研究工作；而傳統被探討的選題方式有：(1)隨機法(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)等五種[9]，其時間複雜度(time complexity) [11]為 $O(m)$ (隨機法)或 $O(mn)$ (其他四種方法) (其中， m 為測驗之題數， n 為題庫中題目數)，執行效率可謂相當高，但由於選題方式過於簡單而缺乏彈性，以致於選出之試題與目標訊息函數間之誤差仍偏高，因此，近年來一些學者，如：Wang & Ackerman[18]與 Swanson & Stocking[16]等人，相繼提出更精準的選題策略，使誤差大量減少，而本研究則擬以類神經網路能量模式[10,15]，設計出一高效率且低誤差的貪婪演算法選題策略，使測驗題目反應出之測驗訊息(test information) 更能符合測驗設計者的需求。第二章，我們將對各種選題策略做一說明，第三章，則介紹本研究所提之貪婪演算法選題策略，第四章則進行各種選題策略之效能評估，最後，做一簡短結論與建議。

2. 各種選題策略之比較

利用題目反應理論來評鑑測驗之優劣的指標不是信度係數，而是測驗訊息[3, 13, 14]。測驗訊息或題目訊息以及能力估計值的標準誤均是能力的函數，而非一個常數，因此，在以題目反應理論為基礎的適性測驗中，每一題目對不同能力(θ)均反應出不同訊息量($I(\theta)$)，而此訊息量將可用來做選題時之參考，當訊息量越大，表示此題目對受試者能力估計越有效，因此，適性測驗可在最短施測題數內估計出受試者之潛

在能力。表一為我們國科會研究計劃中，所建構之國民小學五年級自然科電腦化適性測驗系統中題庫(總題數 160 題)之部份題目(30 題)，題庫建立日期於民國八十七年六月初，施測時間為兩週，經編製雙向細目表、編擬問卷、預試、題目分析、刪題、編擬正式問卷、正式施測、資料整理與分析，歷經二個月。題目為四選一選擇題型式；內部一致性係數(α)為 0.8214；測驗得分和受試者五上及五下自然科學期成績之相關為 0.7001；題庫題目鑑別度介於 0.40 與 2.50 之間；難度位於-3.00 與 3.00 之間；而猜測度則介於 0.08 與 0.44 之間；本研究則以此題庫為例，說明各種選題法運作方法其特性與優缺點。

今以早期所提之五種方法來選 30 題，以建構一單峰型態測驗訊息函數為例，最大峰值在能力為 0，測驗目標訊息函數如表二所示時，各方法之程序分述如下：

(1) 隨機法(random)

使用亂數表，由題庫中任意選取 30 題。

(2) 標準法(standard)

使用傳統題目統計法(题目的難度(p)和鑑別度(r))來選題，首先，選出難度為.30 至.70 之題目，接著再由其中選出鑑別度最大之 30 題。

(3) 中間難度法(middle difficulty)

對所有題目在能力(θ)=0 中，挑出訊息量最大之 30 題。

(4) 上下法(up and down)

此方法包括三個步驟，第一步驟為：在所有題目中，選出題目之能力值(θ)=-1 中，訊息量最大的一題；第二步驟：類似第一步驟，但在題目之能力值(θ)=0 中，選出訊息量最大之一題；第三步驟亦類似，但由題目之能力值(θ)=1 中，選出訊息量最大之一題；此三種選取之題目均不重覆。當此三步驟重覆 10 次，則可選出測驗所需之 30 題題目。

(5) 最大訊息法(maximum information)

由題庫中選出能力值(θ)=-1, 0 和 1 之三個訊息量和，其和最大的 30 題題目。

當採用以上五種方法，自題庫中選取 30 題編制成一份測驗，則最後所求得之測驗訊息量將如圖一所示，圖中明顯表示出傳統五種方法所獲得之測驗訊息量與目標訊息量之間誤差仍偏過大，與早期相關研究者所得之結果類似[8]，表示此五種方法並不十分適用。因此，近年來不少專家學者提出不同數學模式或解題技巧來處理此問題，例如：van der Linden [17] 則以二元規劃(binary programming)問題方式來解，而 Swanson & Stocking [16] 等人採用相似原理，其基本原理為二元線性規劃(binary linear programming)，但為突破線性規劃解題之限制，加入一些啟發式(heuristic)技巧(如：將測驗題目與題庫題目互換(replacement))，使結果得到進一步改善，其選題策略主要包含兩個程序(phase)：程序一(phase 1)，以選題法進行選題(基本上該選題法為一貪婪演算法(greedy algorithm))，先建構一份 m 題測驗，接下來再進行程序二(phase 2)，為一取代程序(replacement phase)，在此程序中，繼續選新的一題，再與測驗中 m 題一一進行互換，看是否有任何一題與此新選的題目互換後，測驗訊息與目標訊息間的誤差可減少，若有，則將減少量最多的一題與新選的題目互換，並由題庫中再選新的一題，重複此互換動作，直到互換動作無法減少誤差為止；當題庫之題目為相同領域時，Swanson 與 Stocking 演算法的運作方式可簡化為：程序一(phase 1) 測驗初步建構

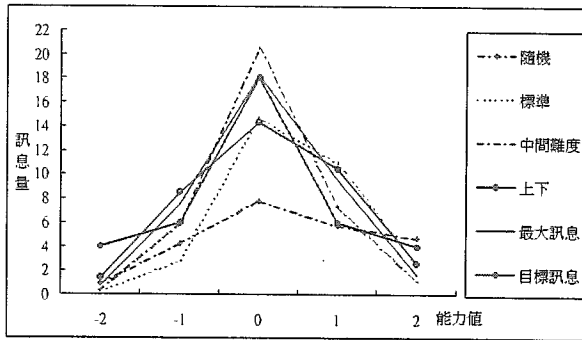
1. 計算題庫中每一題目加入測驗中後，測驗訊息量與目標訊息量之誤差(Swanson and Stocking 方法中有定義測驗訊息函數上下限，本研究中設定為相同)，若滿足所設定之目標訊息函數則停止。
2. 選擇一題目 t ，當加入此題後，可使目標訊息量與測驗訊息量之正誤差最小。
3. 將題目 t 加入測驗中，檢查是否已達到測驗所需題數，若達到則停止選題，進入程序二；否則回步驟 1。

表一、適性測驗題目之題目參數與訊息量及傳統統計法之難度(p)與鑑別度(r)

編號	題目參數			能力等級							傳統統計數	
	a	b	c	-3	-2	-1	0	1	2	3	p	r
1	0.763	-3.000	0.270	0.242	0.193	0.078	0.024	0.007	0.002	0.001	0.969	0.584
2	0.542	-2.218	0.280	0.085	0.123	0.104	0.060	0.028	0.012	0.005	0.897	0.405
3	1.252	0.665	0.330	0.000	0.000	0.007	0.202	0.577	0.155	0.021	0.547	0.442
4	0.648	-2.978	0.250	0.181	0.159	0.081	0.031	0.011	0.004	0.001	0.953	0.455
5	0.756	-1.816	0.260	0.082	0.227	0.215	0.096	0.031	0.009	0.002	0.894	0.607
6	0.445	-1.451	0.290	0.038	0.068	0.082	0.069	0.045	0.025	0.013	0.809	0.299
7	0.446	-1.774	0.270	0.051	0.079	0.084	0.064	0.039	0.021	0.011	0.834	0.322
8	0.732	-3.000	0.320	0.199	0.167	0.073	0.024	0.007	0.002	0.001	0.984	0.245
9	0.647	-0.611	0.360	0.008	0.043	0.119	0.146	0.091	0.038	0.014	0.778	0.386
10	1.594	0.791	0.440	0.000	0.000	0.001	0.081	0.760	0.142	0.010	0.606	0.351
11	0.537	-1.598	0.290	0.049	0.101	0.118	0.085	0.045	0.020	0.009	0.847	0.407
12	0.564	-1.124	0.220	0.038	0.100	0.149	0.126	0.071	0.032	0.013	0.778	0.449
13	0.745	0.989	0.350	0.000	0.001	0.013	0.081	0.194	0.162	0.068	0.534	0.293
14	0.467	0.546	0.360	0.003	0.011	0.032	0.062	0.078	0.066	0.042	0.628	0.228
15	0.493	-1.497	0.290	0.042	0.083	0.100	0.080	0.047	0.024	0.011	0.828	0.333
16	0.601	-2.323	0.270	0.110	0.155	0.116	0.058	0.024	0.009	0.003	0.913	0.471
17	0.728	0.005	0.320	0.002	0.015	0.086	0.197	0.167	0.073	0.024	0.672	0.415
18	0.612	-1.362	0.250	0.045	0.123	0.167	0.120	0.058	0.023	0.008	0.822	0.473
19	0.836	-0.262	0.250	0.002	0.028	0.170	0.312	0.178	0.056	0.014	0.688	0.579
20	0.715	-1.953	0.260	0.097	0.214	0.184	0.083	0.028	0.009	0.003	0.900	0.547
21	0.636	-1.572	0.270	0.055	0.142	0.170	0.106	0.046	0.017	0.006	0.856	0.495
22	0.559	-0.850	0.260	0.023	0.071	0.128	0.128	0.080	0.038	0.016	0.756	0.436
23	1.492	-2.281	0.240	0.235	0.966	0.174	0.015	0.001	0.000	0.000	0.962	0.905
24	0.805	-1.344	0.200	0.044	0.207	0.315	0.172	0.056	0.015	0.004	0.837	0.641
25	0.564	-1.936	0.280	0.073	0.127	0.122	0.074	0.035	0.014	0.006	0.878	0.424
26	1.714	0.924	0.340	0.000	0.000	0.000	0.055	1.088	0.221	0.013	0.488	0.354
27	0.493	-0.998	0.230	0.032	0.074	0.110	0.104	0.069	0.037	0.018	0.753	0.384
28	0.827	-1.629	0.200	0.074	0.276	0.302	0.129	0.037	0.009	0.002	0.875	0.655
29	0.717	-0.985	0.140	0.035	0.150	0.280	0.218	0.095	0.032	0.010	0.759	0.605
30	0.974	-1.803	0.240	0.081	0.381	0.324	0.094	0.020	0.004	0.001	0.913	0.707

表二、測驗目標訊息函數

	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	18	6	4



圖一、測驗目標訊息曲線與五種不同選題策略產生之測驗訊息曲線圖

程序二(phase 2)：取代程序(replacement phase)

4. 以前述程序一選題法，再由題庫選出新的一題。
5. 將測驗中每一題分別與新選的題目互換，看是否有任何一題與此新選的題目互換後，測驗訊息與目標訊息間的誤差可減少，若有，則將減少量最多的一題與新選的題目互換並回步驟 4；若無減少，則停止取代程序，結束此演算法。

此演算法之運算方式與 Wang & Ackerman[18]等人所提方法類似，但加入取代程序(replacement phase (phase 2))，使選題結果之測驗訊息誤差有進一步改善可能，但運算時間複雜度則相同(詳見第四章效能評估)。Wang & Ackerman [18]等人所提之方法，包含 TESTGEN 與 TESTGEN 等演算法，針對不同能力值時題目之訊息量，將題庫中題目排優先(priority)次序，再檢查建構中之測驗與目標測驗最大差距在那一能力值，則從該能力值找出最優先之試題編入測驗中，如此，將可使建構中之測驗與目標測驗差距大幅減少，其中，TESTGEN2 類似 TESTGEN1，但加入考慮不同內容領域訊息，而本研究之選題策略題庫之題目為相同領域，故將只說明 TESTGEN 演算法，其運作方式可簡化為：

1. 任意選取第一題編入測驗中。
2. 計算測驗訊息量。
3. 檢查測驗題數是否已達設定題數(= m)，若是，則停止；否則繼續下一步驟。
4. 計算在不同能力值(θ)時，測驗訊息量與目標訊息量之誤差，並以誤差多寡排列其優先次序(誤差多的優先)。
5. 在第一優先次序之能力值(θ)的條件下，由題庫(題數 = n)中選取一最大訊息量之題目進入測驗中，如此，測驗訊息量最大誤差將可被最

有效減少。

6. 重回步驟 2。

在此演算法步驟 5 中，題庫中的每一題，在每一能力值下，均需比較訊息量大小，以找出最大訊息量題目，所以，需有 n 次比較，而總共需選出 m 題，因此，總運算次數為 $m \times n$ ，有一多項式時間，效率算相當不錯；至於其他如物理學之技術(如：退火(simulated annealing))，也被 Jeng & Shih [12]等人使用於選題策略，可達到十分不錯結果，但此方法之時間複雜度並無一固定上限，為美中不足所在；本文中，我們將提出一貪婪演算法作為選題策略，使建構之測驗更能符合設計者需求。

3. 貪婪演算法選題策略

對選題問題而言，一測驗訊息量之需求，可以使用圖形表示(如圖一曲線“目標訊息”所示)，在每一能力點(θ)，有一對應之訊息量值 d_i ， i 為能力等級指標(如表二所示)，與我們選出題目所形成測驗之訊息量(O_i)，將有一誤差，此時我們可將此誤差定義為一能量函數(energy function)：

$$E_i = \sum_{i=1}^s (d_i - O_i)^2 \quad (1)$$

其中， d_i 為測驗目標訊息量在能力等級指標 i 時之訊息量值， s 為能力等級數(例如：我們只探討能力值為-2.0, -1.0, 0.0, 1.0, 2.0 時，能力等級數 s 為 5)， O_i 為以選題策略產生出之測驗訊息量值。愈小的 E_i 值(Error, 訊息量誤差值)，則愈符合設計者所需求之測驗目標。另外，以選題策略產生之訊息量 O_i 如下式(2)所示：

$$O_i = \sum_{j=1}^n w_{ij} x_j, x_j \in \{0,1\} \quad (2)$$

其中， w_{ij} 表示題目 j 在能力等級指標 i 時之訊息量； x_j 表示題庫第 j 題題目是否被選擇放入測驗試題中，若放入測驗中則為 1，否則為 0； n 為題庫之總題數。能量函數 E_i 可經由選取題目編入測驗而改變，當題目 j 被選取時，則 $x_j=1$ ，可得能量函數修正量為：

$$\Delta E_{i,j} = E_{i,j} - E_i, \quad (3)$$

其中

$$E_{i,j} = \sum_{i=1}^s (d_i - O_i')^2, \quad (4)$$

與

$$O_i' = \sum_{\substack{k=1 \\ k \neq j}}^n w_{ik} x_k + w_{ij} \cdot 1 = O_i + w_{ij} \quad (5)$$

式(5)代入式(3)中，能量函數修正值為：

$$\begin{aligned} \Delta E_{i,j} &= \sum_{i=1}^s (d_i - O_i')^2 - \sum_{i=1}^s (d_i - O_i)^2 \\ &= \sum_{i=1}^s (d_i^2 - 2d_i O_i' + (O_i')^2 - d_i^2 + 2d_i O_i - O_i^2) \\ &= \sum_{i=1}^s (-2d_i w_{ij} + 2O_i w_{ij} + w_{ij}^2) \\ &= \sum_{i=1}^s w_{ij} (w_{ij} + 2O_i - 2d_i) \end{aligned} \quad (6)$$

當 x_j 變為 1 後，新的訊息函數低於目標訊息函數時，則能量函數 E_I 修正值小於 0，表示能量繼續下降中，建構中之測驗訊息函數逐漸逼近目標訊息函數，在以此觀念為基礎下，發展出本研究之貪婪演算法：

貪婪演算法之選題策略

由於所採用的技術為一平行處理方法，每一變數 x_j 之修正為互相獨立，但此時若同時修正多個變數，將造成能量函數上下震盪無法收斂問題；因此，在每一次迴圈 (iteration) 中，則以改變一個“最適合”之變數 x_j 為原則(將於下列步驟五中說明)，將其狀態改為 1，直到能量函數達到一穩定狀態或題數上限為止。詳細步驟說明如下：

步驟一、設定所有變數初始值

設定： d_i 為能力值指標 i 之測驗目標訊息量值($i=1 \sim 5$)；題庫中所有題目 j ($j=1 \sim n$)之選取狀態 x_j 為 0 (未選取入測驗中)； w_{ij} 為每一題目 j 在能力指標 i 之訊息量值；初始時間變數 $t=0$ 。

步驟二、求測驗訊息函數 $O_i(t)$ 。

$$O_i(t) = \sum_{j=1}^n w_{ij} x_j(t), \quad \forall i = 1 \sim s. \quad (7)$$

步驟三、求誤差函數(能量函數)

如前式(1)，再加入運算迴圈次數 t ：

$$E_I(t) = \sum_{i=1}^s (d_i - O_i(t))^2 \quad (8)$$

步驟四、求每一題目 j 選取後，能量函數修正值

$$\Delta E_{I,j}(t):$$

$$\Delta E_{I,j}(t) = \sum_{i=1}^s w_{ij} (w_{ij} + 2O_i(t) - 2d_i) \quad (9)$$

步驟五、選擇一“最適合”的變數作狀態改變

由所有 $x_j \neq 1$ 中，找一能量函數 E_I 修正量

最小值，即

$$\begin{aligned} \Delta E_{I,j^*}(t) \\ = \text{Min}\{\Delta E_{I,j}(t), \forall x_j(t) \neq 1, j = 1 \sim n\} \end{aligned} \quad (10)$$

步驟六、修改變數 x_j^* 之狀態為 1。

$$x_j^*(t+1) = 1 \quad (11)$$

步驟七、計算能量函數之改變量

$$\Delta E_I(t+1) = E_I(t+1) - E_I(t). \quad (12)$$

步驟八、檢查是否收斂至穩定值

(1) 無固定題數限制

此時須檢查能量(誤差)是否一直遞減(<0)，直到無法遞減(≥ 0)為止

$$\Delta E(t+1) \geq 0 \quad (13)$$

(若加入題數上限(= m)時，則加入題數限制之檢查，如下列條件(2))

(2) 有固定題數限制

此時須比較選取的題數是否為 m 題(或上限為 m 題)：

$$\left(\sum_{j=1}^n x_j\right) - m = 0. \quad (14)$$

(1)或(2)中若有答案成立時，則至步驟八；

否則， $t = t+1$ ，回至步驟二。

步驟八、結束

此時，對所有 $x_j = 1$ 的題目，即為被選取而編入測驗之試題。

此演算法可平行計算，但每次只修改單一個變數狀態，使建構之測驗能逐漸逼近目標訊息函數，以編制一份最符合測驗訊息量需求之測驗，下一章節將進行效能評估與時間複雜度分析。

4. 效能評估

我們以先期研究完成之適性測驗題庫為依據，進行傳統五種選題法與本研究法效能之比較，其結果如圖二(無固定題數限制，但上限為 35 題)與圖三(固定題數 30 題)

所示，由兩圖之結果顯示，我們所提的方法其效能遠優於早期所提之方法(如表三所示)，尤其在固定題數限制下，更顯示出本方法之優點，證明了本研究之價值。另外，本研究法所須的時間，我們亦進行分析與比較，表四為傳統五種方法與本研究法之時間複雜度，我們的方法與其他方法的時間複雜度相同，並未因效能的提升而

需付出大量時間的代價。此外，我們亦與近期學者 Wang & Ackerman [18]及 Swanson & Stocking[16]等人之方法做比較，經實際模擬選題後，結果如圖四與圖五所示，其誤差平方和如表五所示，我們的結果亦優於新近研究方法。

表三、貪婪演算法選題策略與傳統五種選題法之誤差量平方和(SE)

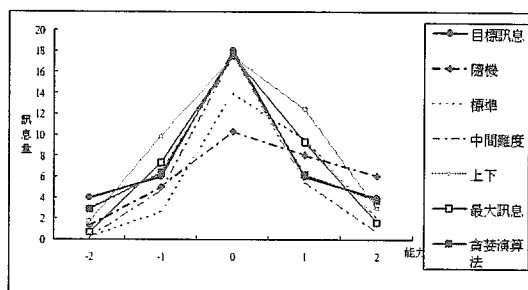
	貪婪演算法	隨機法	標準法	中間難度法	上下法	最大訊息法
圖二(無固定題數，但上限為 35 題)	2.773	75.678	57.662	27.712	46.787	29.591
圖三(固定題數 30 題)	4.504	130.528	61.952	30.867	63.355	31.012

表四、傳統五種選題策略與貪婪演算法之時間複雜度 (m ：表測驗之題數， n ：表題庫之題數($m \ll n$))

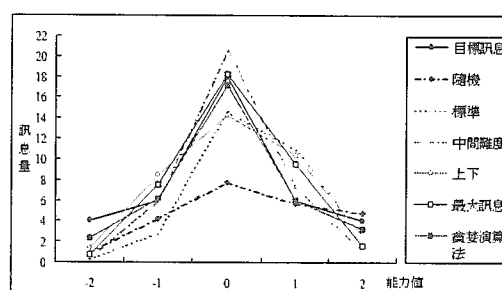
方法	隨機法	標準法	中間難度法	上下法	最大訊息法	貪婪演算法	Wang & Ackerman	Swanson & Stocking
時間複雜度	$O(m)$	$O(mn)$	$O(mn)$	$O(mn)$	$O(mn)$	$O(mn)$	$O(n \log n + mn)$	$O(mn)$

表五、貪婪演算法選題策略、Wang & Ackerman 及 Swanson & Stocking 等方法之誤差量平方和(SE)

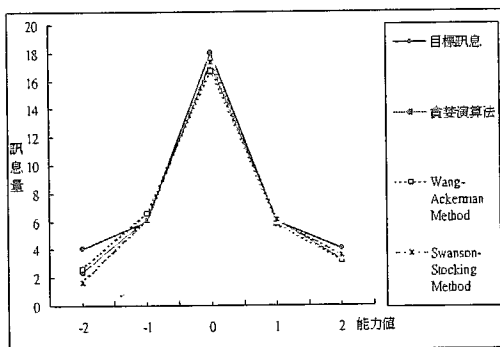
	類神經網路法	Wang & Ackerman	Swanson & Stocking	
			取代程序前	取代程序後
圖四(無固定題數，但上限為 35 題)	2.773	3.799	3.885	3.087
圖五(固定題數 30 題)	4.504	4.783	6.081	6.081



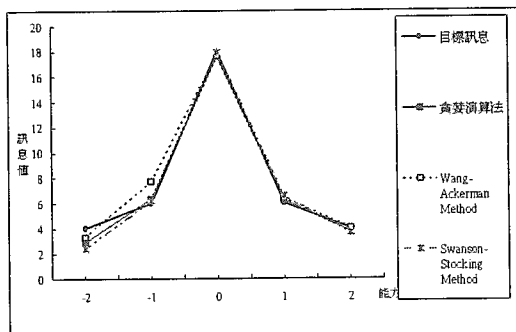
圖二、貪婪演算法選題策略與傳統選題法在無固定題數(上限為 35 題)條件下之測驗訊息函數曲線圖



圖三、貪婪演算法選題策略與傳統選題在固定題數(30 題)條件下之測驗訊息函數曲線圖



圖四、貪婪演算法選題策略、Wang & Ackerman 及 Swanson & Stocking 等方法在無固定題數(上限為 35 題)條件下之測驗訊息函數曲線圖



圖五、貪婪演算法選題策略、Wang & Ackerman 及 Swanson & Stocking 等方法在固定題數(30 題)條件下之測驗訊息函數曲線圖

在分析時間複雜度上，我們分成兩種狀況：(1)題數不限定(但上限為 m 題)，(2)測驗題數限定為 m 題；因此，傳統的標準法、中間難度法、上下法、最大訊息法等(隨機法不列入分析)，不論在狀況(1)或(2)條件下，每次(每一題)均需由題庫中找最大值(時間複雜度為 $O(n)$)，最多找 m 題(次)，所以，總複雜度為 $O(mn)$ ；而我們所提的貪婪演算法選題策略，每次題目選取時，需計算修改每一 x_j 狀態後，能量函數之修正值，為一固定計算時間 $O(s)$ (s 為一常數值，如式(9)所示)，而題庫中共有 n 個 x_j 須計算，所以，每選一題所須之計算亦為 $O(n)$ ，最多找 m 題，所以，總複雜度也為 $O(mn)$ ，與傳統選題策略相同，並未增加計算之複雜度；而 Wang & Ackerman 方法對題目在不同能力值須先排優先次序，其時間複雜度為 $O(n \log n)$ ，隨後則依其演算法進行計算、比較與選題，須 $O(mn)$ 之計算，所以時間總複雜度為 $O(n \log n + mn)$ ；Swanson & Stockin 方法在程序一(phase 1)為計算測驗訊息誤差並選擇一誤差最小的題目(類似貪婪

(greedy)演算法)，須 $O(mn)$ 處理時間，隨後之測驗與題庫中題目互換程序(phase 2)，其時間上限亦為 $O(mn)$ ，所以時間總複雜度仍為 $O(mn)$ ；這兩種新近提出的選題方法，其時間複雜度均與我們相近或相同，所得結果則略遜於我們，此外，本文所提之方法，更具一般性與彈性，因只需考量如何降低一數學函數之能量值即可。

5. 結論與建議

本研究中，我們以自行設計的貪婪演算法作為選題策略，以此技術建構出之測驗其訊息函數與測驗目標之訊息函數，兩者間的誤差遠低於早期研究的五種選題策略(隨機法，標準法，中間難度法，上下法及最大訊息法如前表三所示))，並優於近年來專家學者們(如 Wang, Ackerman, Swanson 與 Stocking 等人)之研究結果(如前表五所示)，但是我們並未因此而增加時間複雜度。由於，本文所提方法是將選題問題變為一數學函數，只需在符合選題問題之限制下，求得此數學函數之最佳解或接近最佳解，即可建構一份最接近目標訊息量之測驗，因此，本研究擴展測驗問題之呈現方式與解決方法，使測驗目標更能符合設計者之需求，對教育評量研究者提供另一有效的選題工具；我們將進一步研究其它資訊技術，如何運用於教育研究相關領域中，以提早教育科技品質與水準，使教育工作達到事半功倍之效果。

誌謝

本研究計畫經費蒙國科會補助(編號 NSC 88-2511-S-024-005-)，謹此致謝。

參考文獻

- [1] 何榮桂。電腦化測驗概述。現代教育，第 18 期，121-129 頁，民 80。
- [2] 何榮桂、杜玲均、莊謙本。改良式之灰色預測電腦化適性測驗選題策略。第七屆國際電腦輔助教學研討會大會論文，393-400 頁，高雄，國立高雄師範大學，3 月 19~21 日，民 87。
- [3] 吳裕益。電腦化適性測驗與傳統測驗之比較。教師天地，第 54 期，民 80 年 10 月，49-53 頁，民 80。
- [4] 杜淑芬、黃國禎。網路測驗及評估系統試題動態配置最佳化之研究。第七屆國際電腦輔助教學研討會大會論文，337-343 頁，高雄，國立高雄師範大學，

- 3月19~21日, 民87。
- [5] 孫光天、陳新豐、吳鐵雄。線上適性測驗回饋對作答情緒與動機影響之研究。第七屆國際電腦輔助教學研討會學生論文專題論文集編, 9-14頁, 高雄, 國立高雄師範大學, 3月19~21日, 民87。
- [6] 孫光天、鄭海東、謝凱隆、陳新豐。智慧型線上適性測驗系統。第七屆國際電腦輔助教學研討會學生論文專題論文集編, 81-86頁, 高雄, 國立高雄師範大學, 3月19~21日, 民87。
- [7] 孫光天、陳新豐、楊振印、戴伯昌。線上適性整體評量環境之研究。TANET'98, 1A13-22頁, 花蓮, 國立東華大學, 11月9~11日, 民87。
- [8] Cool, L. L., & Hambleton, R. K. A comparative study of item selection methods utilizing latent trait theoretic models and concepts, Laboratory of Psychometric and Evaluative Research Report, No. 88. Amherst: University of Massachusetts, School of Education, 1978.
- [9] Hambleton, R. K., & Swaminathan, H. ITEM RESPONSE THEORY-- Principles and Applications, Kluwer Academic Publishers Group, Netherlands, 1985.
- [10] Hopfield, J. J., & Tank, D. W. "Neural computation of decision in optimization problems," *Biological Cybernetics*, 52, 141-152, 1985.
- [11] Horowitz, E., & Sahni, S. *Fundamentals of Computer Algorithms*, Computer Science Press, Potomac, MD, 1978.
- [12] Jeng, H. L., & Shih, S. G. "A Comparison of Pairwise and Group Selections of Items Using Simulated Annealing in Automated Construction of Parallel Tests," *Psychological Testing*, 44 (2), 195-212, 1997.
- [13] Lord, F. M., & Novick, M. R. *Statistical theories of mental test scores*, Reading Mass: Addison-Wesley, 1968.
- [14] Lord, F. M. *Applications of item response theory to practical testing problems*, Hillsdale, NJ: Erlbaum, 1980.
- [15] Sun, K. T., & Fu, H. C. "A Hybrid Neural Network Model for Solving Optimization Problems," *IEEE Transactions on Computers*, 42(2), 218-227, 1993.
- [16] Swanson, L. & Stocking, M. L. "A model and heuristic for solving very large item selection problems," *Applied Psychological Measurement*, 17, 151-166, 1993.
- [17] van der Linden, W. J. "Automated test construction using minimax programming," In W. J. van der Linden (Ed.) *IRT-based test construction* (pp. 1-16). Enschede, The Netherlands: Department of Education, University of Twente, 1987.
- [18] Wang, C. S. & Ackerman, T. "Two Item Selection Algorithms for Creating Weakly Parallel Test Forms Using the IRT Information Functions," *Psychologica Testing*, 44(2), 123-140, 1997.