

# 台北市長選舉之基因預測模型

劉文卿\*

## 《本文摘要》

本論文引用遺傳演算法的觀念，建立基因模型，以預測選舉結果，其基本假設是：每一位民意調查的受訪者均有一選舉染色體，該染色體決定了受訪者的投票行爲，受訪者的選舉染色體利用問卷建立，並進而依據受訪者的染色體建立候選人的選舉染色體，受訪者會將其選票投給與其染色體最類似的候選人。利用基因模型，對八十三年台北市長選舉實施預測，可達到低於1%的預測精度。

## 壹、前言

選舉預測是研究選舉的學者關心的問題，相關的學術著作頗多，張紘炬(1986)以行政區、年齡、性別、籍貫、教育程度、黨籍等因素，分別預測各候選人單項因素的獲票率，再將各因素整合，求得綜合獲票率。陳義彥(1989)以省籍、年齡、教育程度、收入、政黨偏好、對政府的信任程度、對政府的評價及政治知識等八個變數，實施區別分析，建立群組，並進而將未回答將投票給何人者歸入群組以預測投票結果。陳義彥(1994)進一步將陳義彥(1989)的觀念定型化，提出群集分析的觀念，以政黨、候選人、政見、族群等四個因素實施集群分析，並依據集群特性來預測政黨所獲得的支持率。洪永泰(1994)的預測方法分兩階段：首先，以投票所的歷年投票記錄為基礎，將各投票所分類成爲集群，並建立各集群的投票傾向，其次將民意調查中未表示投票對象者，歸入各集群，並依集群傾向預測其投票對象，洪永泰(1994)與陳義彥(1994)的主要差別有二：1. 洪永泰

---

\* 國立政治大學資訊管理學系

以村里為集群分析的基礎，而陳義彥是以民意調查為基礎；2. 洪永泰的集群分析在不同縣市採用不同變數，以符合區域特性，而陳義彥是採用一致的變數；換言之洪永泰的集群分析以行政區域為基礎，而陳義彥則無此項考慮。吳統雄(1994)利用受訪者對候選人的評價(0-100)，計算候選人的形象指標及形象動力指標，並分析指標高低與當選、落選間的關係。

本論文採取不同的預測模式，引用遺傳演算法 (genetic algorithm, GA) 的觀念，建立基因模型，以預測選舉結果，其基本假設是：每一位民意調查的受訪者均有一選舉染色體，該染色體決定了受訪者的投票行為，受訪者的選舉染色體利用問卷建立，並依據受訪者的染色體建立候選人的選舉染色體，投票人會將其選票投給與其染色體最類似的候選人。本文所提出的基因模型與陳義彥的集群分析，雖在外觀上不同，但基本精神仍可相通，詳細討論見第四節。利用基因模型除了可以進行預測外，尚可找出問卷中對預測精度影響較大的問題（在本論文中稱為解析度），供作問卷設計的參考。

## 貳、資料來源

本研究的資料來源是政治大學選舉研究中心陳義彥教授主持的國科會專題研究計畫「台北市長選舉預測模型」。在此計畫中選前共做了兩次民意調查，選後做了一次民意調查，本研究引用第一次及第二次選前民意調查的資料進行預測，並與實際的選舉結果比較。

## 參、基因模型

本論文採用基因模型作為預測的方法，基因模型的最大特色是問卷中所有與投票相關的問題及選項均列入模型中，並在預測的時候發生一定的影響力。基因模型的原始觀念來自遺傳演算法(genetic algorithm) (Goldberg, 1989)，遺傳演算法係一種隨機搜尋的方法，以搜尋最佳解為目標。在本論文中僅引用其基因結構的觀念，而並不實施字串運算、模式比對、及隨機搜尋等動作，故不可視為以遺傳演算法來解決選舉預測的問題。

### 一、基因、基因組與染色體

基因模型的基本假設是：受訪者的投票行為由他的選舉染色體(election chromosome)決定，選舉染色體利用受訪者所填的問卷來建立。

一份問卷由若干問題組成，每一個問題由若干選項組成，在基因模型中，每一個選項代表一個基因(gene)，每一個問題代表一個基因組，問卷代表一個染色體

(chromosome)，受訪者填完一份問卷，便是建立了他的選舉染色體，顯示了他在受訪時的投票傾向。一個受訪者在不同的時間對同一問卷可能會有不同的答案(建立不同的染色體)，說明投票傾向的改變。同一問題的選項合成一基因組，同一基因組的基因代表受訪者投票傾向可能的變化，被選中的基因稱為優勢(dominance)基因，未被選中的基因稱為擱置(abeyance)基因，受訪者的投票傾向由其全部的優勢基因共同決定，優勢基因的選擇改變，投票傾向也就變了。隨著外在條件的改變，一個基因組的優勢基因，可能會由同一組中的擱置基因替代。例如，假設在兩岸關係的問題上，有獨立與統一兩個選項，這兩個選項合成一基因組，受訪者若主張統一，則是以統一為優勢基因，而以獨立為擱置基因，隨著兩岸環境的改變，同一受訪者可能選擇獨立為優勢基因，而以統一為擱置基因。優勢基因與擱置基因是生物體適應外在環境變動的重要機構，在選舉染色體上也可反映受訪者投票傾向的變化。受訪者的投票傾向由其全部的優勢基因共同決定，其中影響力最大的基因所屬的基因組稱為優勢基因組，例如在問卷中直接詢問受訪者會投票給誰，便可視為優勢基因組。

在本研究中，問卷視為染色體結構，每一問題 $i$ 由一基因組 $g_i$ 表示，問卷若包含 $m$ 個問題，則該問卷染色體 $C$ 由 $m$ 個基因組 $g_i$ 組成，如下式

$$C = g_1 g_2 \dots g_m$$

若一問題包含 $d$ 個選項，則該問題之基因組 $g_i$ 包含 $d$ 個位元，如下式

$$g_i = b_1 b_2 \dots b_d$$

一受訪者在填寫問卷之前，所有基因的位元值均為0，當對問題 $i$ 設定選項後，基因組 $g_i$ 中相對於該選項的基因值改為1，例如受訪者選擇2，則 $g_i = 010\dots 0$ ，若一問題為單選，則基因組內只能有一個基因為1，若為複選，則可有多個基因為1。在問卷全部填完後，該受訪者的染色體便完成了。

在此架構下，一問卷染色體 $C$ 是由0與1組成的字串(string)，若有 $m$ 個問題且第 $i$ 個問題有 $d_i$ 個選項，則 $C$ 的字串總長度為 $n = \sum_{i=1}^m d_i$ ， $C$ 可以下式代表

$$C = b_1 b_2 b_3 \dots b_n$$

以下除非另有說明，問卷染色體 $C$ 均視為由 $n$ 個0與1位元組成的字串，此 $n$ 個位元分成 $m$ 個基因組。

## 二、候選人染色體

利用受訪者染色體可建立候選人的選舉染色體，其方法如下：將支持某一候選人的所有受訪者的染色體相加後求其平均值，則可得出候選人的染色體。此一定義所得的候選人的選舉染色體係由其支持者的投票傾向共同決定。候選人染色體的結構與受訪者有一點不同：候選人的基因值因為由平均而得，故是介於0與1之間的實數，而非單純的0或1。由上述可知一候選人k的染色體 $C_k$ 包含n個實數：

$$C_k = f_1 f_2 \dots f_n$$

$f_i$ 計算如下：

$$f_i = \frac{\sum_s b_i}{s} \quad (1)$$

在(1)式中， $f_i$ 是候選人k的第i個基因值，s代表受訪者中支持k的人數， $\sum_s b_i$ 代表s個支持者的染色體在第i個基因（位元）的和。換言之， $f_i$ 的值等於全部支持者第i個基因的和均值(average sum)，因為是平均值，故介於0到1之間。在一基因組內，若支持者全部以某一基因為優勢基因，則候選人在該基因的值等於1，若全部以該基因為擱置基因，則候選人該基因的值為0，多數情況下不會等於0或1，當一基因值大於基因組內的平均基因值時稱為優勢基因。若基因組為單選，同一基因組內的基因和仍為1。

## 三、預測力與解析度

預測力是指一個選舉模型，對選舉結果預測的準確程度，本論文以預測得票率與實際得票率的平均差為預測力的評斷標準，定義如下：

$$p = \frac{\sum_{i=1}^{n_c} |v_i^a - v_i^p|}{n_c} \quad (2)$$

(2)式中p為預測力， $n_c$ 為候選人的總人數， $v_i^a$ 與 $v_i^p$ 分別為第i個候選人的實際得票率及預測得票率。

解析度(resolution)有兩種：基因解析度與基因組解析度。基因 $f_i$ 的解析度 $r_i$ 是指兩兩候選人在此基因的差異和(sum of differences)，以台北市長選舉為例，可兩兩配對合成三組，三組在 $f_i$ 的差值的和稱為該基因的解析度，若以(a,b)代表兩兩配對則

$$r_i = \sum |f_i^a - f_i^b|, \quad (a,b)$$

差異和的值越大，表示候選人在此基因的差異性越大，受訪者越容易就此基因區別出其投票傾向。因為候選人的基因值在0與1之間，故一個候選人的值為1，而別人均為0時，解析度最高，因此解析度的值為 $0 \leq r_i \leq n_c - 1$ ，其中 $n_c$ 代表候選人數。若全部候選人在某一基因的值均很接近，則 $r_i \rightarrow 0$ 表示該基因的解析度很低，反之該基因具有高解析度。基因組解析度是該基因組內各基因解析度的和。台北市長候選人雖有四位，但本研究僅考慮黃大洲、陳水扁、趙少康等三位，故 $n_c=3$ ，因此基因解析度 $r_i$ 的上限為2，在同一基因組內，最多同時有三個基因達到上限值，因此基因組解析度的最大值為6。

#### 四、預測方式

候選人染色體代表該候選人對受訪者的吸引力，候選人染色體  $C=f_1f_2\dots f_n$  及受訪者染色體  $V=b_1b_2\dots b_n$ ，兩者間吸引力的計算方式如下：

$$F = \sum_{i=1}^n f_i * b_i$$

(3)中 $0 \leq f_i \leq 1$ ， $b_i = \begin{cases} 1 \\ 0 \end{cases}$ 。若候選人和受訪者均以某一基因 $i$ 為優勢基因，則 $f_i * b_i \rightarrow 1$ ，若候選人和受訪者有多個相同的優勢基因，則 $F \rightarrow n$ ，吸引力的值介於0與 $n$ 之間。不同的候選人對受訪者的吸引力會不同，本論文假設受訪者會將選票投給吸引力最大的候選人。在基因模型中，因為所有選項均列為基因，支持者與候選人間，縱使有相當的差異存在，但只要兩者間的吸引力大於其他候選人，仍可正確的判斷支持者與被支持者間的投票關連。

由實際資料顯示，部份受訪者雖明白表示支持某一候選人，但兩者間的吸引力，不及其他候選人，本論文稱之為雜訊染色體。雜訊染色體是指一個染色體所顯示的投票傾向與其表態支持的對象不同。雜訊染色體的辨認很單純，從吸引力的大小，便可判斷，消除雜訊染色體的方法是在建立候選人染色體之後，不論受訪者是否表態，全部重新計算其吸引力。

## 肆、實驗結果

### 一、前處理

實驗數據是利用選研中心在選舉前所做的兩次民意調查，第一次民意調查由八十三年十月二十八日至十一月五日，有效樣本共1044份；第二次民意調查由八十三年十一月

二十二日至十二月一日，有效樣本共1097份。第一次選前民意調查共有六十六個問題，第二次調查共有五十七個問題，兩次調查中若干問題與受訪者的投票意向完全無關，例如：電話號碼尾數、撥號次數、成年男性幾人、請某某人聽電話等等，這些問題在實驗中均予刪除，刪除後，第一次調查尚餘五十八個問題，第二次調查尚餘四十二個問題。每一個問題的選項數目不同，最少的有兩個選項（性別），最多的有八十七個項（台北市最需要解決的問題）。中性選項如：很難說、不知道、無意見、拒答等，在計算與候選人間的吸引力時，一概以0計算。

可用來作為優勢基因組的問題有三個：「比較希望那一個當選」、「明天投票會把票投給誰」、「比較可能投給誰」。本論文選擇「明天投票會把票投給誰」這一問題作為優勢基因組。「明天投票會把票投給誰」共有1.黃大洲、2.陳水扁、3.趙少康、4.黃+陳、5.陳+趙、6.黃+趙、7.紀榮治、92.不去投票、93.投廢票、95.拒答、96.尚未決定、97.無意見、98.不知道等十三個選項。選擇1、2、3三個選項的受訪者列為原始支持者，其餘均視為隱藏支持者，其數量如下表所示。

為求簡化問題，表一與表二中的實際得票率，是以黃大洲、陳水扁、趙少康等三人的得票數來計算，紀榮治的得票數和廢票數未列入考慮。第一次調查1044位受訪者中，原始支持者共有546位，隱藏支持者共498位，原始支持率與實際得票率的誤差為8.83%，預測力為2.94%。第二次調查1097位受訪者中，原始支持者共有599位，隱藏支持者

表一 選前第一次民意調查的原始預測力

	原始支持數	原始支持率	實際得票率	誤差	原始預測力
黃大洲	144	26.37%	25.96%	0.41%	
陳水扁	215	39.38%	43.79%	4.41%	
趙少康	187	34.25%	30.25%	4.00%	
總計	546	100.00%	100.00%	8.83%	2.94%

表二 選前第二次民意調查的原始預測力

	原始支持數	原始支持率	實際得票率	誤差	原始預測力
黃大洲	213	35.56%	25.96%	9.60%	
陳水扁	216	36.06%	43.79%	7.73%	
趙少康	170	28.38%	30.25%	1.87%	
總計	599	100.00%	100.00%	19.20%	6.40%

共498位，原始支持率與實際得票率的誤差為19.20%，預測力為6.40%。兩次問卷的內容幾乎雷同，但預測力相差兩倍有餘，所隱涵的意義，值得在日後設計問卷時加以注意。

## 二、預測結果

預測結果分兩部份討論：預測力與解析度。

### (一) 預測力

預測方式依兩種模式進行，第一種模式僅預測隱藏支持者的支持對象，簡稱隱藏模式，第二種模式則預測所有受訪者（不論是否表態）的支持對象，簡稱全體模式。

#### 1. 隱藏模式的預測力

首先依據各候選人的原始支持者的染色體建立候選人染色體，再計算候選人與隱藏支持者間的吸引力，並假設隱藏支持者會將選票投給吸引力大的候選人。依據選前兩次調查資料，所得到的預測結果如下：

表三 第一次民意調查隱藏模式預測力

	原始支持數	預測隱藏支持數	預測總支持數	預測支持率	實際得票率	預測誤差	預測力
黃大洲	144	114	258	24.71%	25.96%	1.25%	
陳水扁	215	244	459	43.97%	43.79%	0.18%	
趙少康	187	140	327	31.32%	30.25%	1.07%	
總計	546	498	1044	100.00%	100.00%	2.49%	0.83%

表四 第二次民意調查隱藏模式預測力

	原始支持數	預測隱藏支持數	預測總支持數	預測支持率	實際得票率	預測誤差	預測力
黃大洲	213	114	327	29.81%	25.96%	3.85%	
陳水扁	216	258	474	43.21%	43.79%	0.58%	
趙少康	170	126	296	26.98%	30.25%	3.27%	
總計	599	498	1097	100.00%	100.00%	7.70%	2.57%

## 2. 全體模式的預測力

隱藏模式假設受訪者表態的支持對象不會改變，但實際上部份受訪者雖然表示支持某一候選人，但因另一候選人的吸引力較大，受訪者極有可能在投票時改變心意另投他人，若考慮此一因素，應在候選人染色體建立後，重行計算對每一受訪者的吸引力，並預測全部受訪者的投票對象。第一及第二次選前民意調查全部重算的結果如下。

表五 第一次民意調查全體模式預測力

	全部重算 支持數	全部重算 支持率	實際得票 率	預測誤差	預測力
黃大洲	247	23.66%	25.96%	2.30%	
陳水扁	464	44.44%	43.79%	0.65%	
趙少康	333	31.90%	30.25%	1.65%	
總計	1044	100.00%	100.00%	4.60%	1.53%

表六 第二次民意調查全體模式預測力

	全部重算 支持數	全部重算 支持率	實際得票 率	預測誤差	預測力
黃大洲	301	27.44%	25.96%	1.48%	
陳水扁	488	44.48%	43.79%	0.69%	
趙少康	308	28.08%	30.25%	2.17%	
總計	1097	100.00%	100.00%	4.35%	1.45%

### (二) 解析度

解析度是指一個問題能鑑別不同候選人的吸引力的程度，兩次民意調查各問題的解析度如表七所示，表七中共有五十八個問題，部份問題在第二次民意調查中並未列入故為空白，問題的次序依第一次民意調查的解析度順序排列，除解析度外，尚列出表達意見的受訪人數及平均解析度，平均解析度因為值太小故乘以100倍。



表七之一

	問 題	解析度	人數	平均	解析度	人數	平均
		第一次民意調查			第二次民意調查		
1	“明天投票會把票投給誰”	6	551	1.088929	6	610	0.983607
2	“比較希望那一位當選”	5.248798	666	0.788108	5.587436	815	0.685575
3	“國民黨、民進黨、新黨較支持哪一黨”	3.877197	853	0.454537	4.152478	961	0.4321
4	“最能了解民眾的需要”	3.21244	674	0.746623	3.087384	708	0.436071
5	“陳水扁、趙少康來比的話，誰會做的比較好”	3.15153	711	0.443253	3.227314	709	0.455192
6	“最有魄力”	3.104762	723	0.429428	3.604118	707	0.509776
7	“做事的能力最強”	3.070358	680	0.451523	3.583201	726	0.493554
8	“對這些問題比較同意那位候選人主張”	2.985991	515	0.579804	3.502516	625	0.560403
9	“最清廉不貪汙”	2.464812	586	0.420616	2.808578	584	0.480921
10	“對黃大洲做市長表現滿意不滿意”	2.45808	882	0.278694	2.5928	964	0.268963
11	“支持不支持民進黨”	2.330904	783	0.297689			
12	“趙少康做他會不會做得比黃大洲好”	2.235197	619	0.361098	2.388677	618	0.386517
13	“這次辯論哪位候選人的表現最好”	2.186845	443	0.493644			
14	“支持不支持國民黨”	2.180967	920	0.237062			
15	“支持不支持新黨”	1.986639	720	0.275922			
16	“陳水扁做他會不會做得比黃大洲好”	19.40528	511	0.379751	2.209241	529	0.417626
17	“民國哪一年出生的”	1.813558	1023	0.177278	1.820499	1069	0.170299
18	“籍貫”	1.69954	1026	0.165647	1.970588	1065	0.185032
19	“對統獨問題的看法”	1.461402	937	0.154966	2.003467	949	0.211114
20	“台灣人、中國人或都是”	1.339585	997	0.134362	1.504727	1058	0.142224
21	“最高學歷”	1.137523	1038	0.109588	0.961146	1080	0.088995
22	“職業”	1.076754	1036	0.103934	0.927443	1089	0.085165

表七之二

23	“投票時是看政黨、候選人還是政見”	0.986103	941	0.104793	1.30017	998	0.130278
24	“台北市最需要解決的問題有二”	0.92066	706	0.130405	1.118391	761	0.146963
25	“最喜歡看那一台的電視新聞”	0.485965	1038	0.0815	1.248248	1010	0.123589
26	“戶籍是在那一區”	0.842624	1035	0.081413	1.154477	1084	0.106502
27	“較偏國民黨、民進黨，還是新黨”	0.619505	354	0.175001	0.686948	403	0.170459
28	“台北市最需要解決的問題有三”	0.610084	323	0.188881	0.805819	344	0.23425
29	“大安區那一里”	0.484418	99	0.489311	0.107756	36	0.299321
30	“81年立委選舉有沒有投票”	0.428024	1044	0.040998	0.379519	1084	0.035011
31	“台北市最需要解決的問題有一”	0.426076	957	0.044522	0.835656	1019	0.082007
32	“士林區那一里”	0.425391	106	0.401312	0.418555	94	0.445272
33	“北投區那一里”	0.379267	98	0.387007	0.322544	91	0.354444
34	“文山區那一里”	0.379009	92	0.411966	0.361237	91	0.396964
35	“信義區那一里”	0.352345	70	0.50335	0.356226	77	0.462631
36	“替誰拉票一”	0.315714	168	0.187925			
37	“內湖區那一里”	0.299127	69	0.433518	0.407905	88	0.463528
38	“哪些人來拉票一” “有沒有看台北市市	0.285408	177	0.161248			
39	長候選人的電視辯論”	0.28268	1044	0.027077			
40	“萬華區那一里”	0.274285	61	0.449647	0.332566	68	0.489068
41	“松山區那一里”	0.25568	56	0.456572	0.232046	54	0.429715
42	“使用的語言”	0.250735	1044	0.024017	0.382353	1097	0.034854
43	“中正區那一里”	0.215814	38	0.567933	0.273091	55	0.496528
44	“中山區那一里”	0.206233	44	0.468712	0.194899	54	0.360923
45	“大同區那一里”	0.127765	28	0.456305	0.212394	42	0.505699
46	“南港區那一里”	0.118742	34	0.349241	0.175989	38	0.463129
47	“性別”	0.112532	1044	0.010779	0.346405	1097	0.031578

表七之三

48	“哪些人來拉票二”	0.095143	28	0.339795			
49	“會不會去投票”	0.069743	983	0.007095	0.037559	1059	0.003547
50	“替誰拉票二”	0.052212	40	0.130531			
51	“哪些人來拉票三”	0.041667	12	0.347222			
52	“訪問結果”	0.0391	92	0.0425			
53	“有沒有人來拉票”	0.031654	1042	0.003038			
54	“哪些人來拉票四”	0.013889	3	0.462963			
55	“哪些人來拉票五”	0.013889	2	0.694444			
56	“比較可能投給誰”	0.009302	100	0.009302	0.011765	90	0.013072
57	“替誰拉票三”	0.007909	9	0.087883			
58	“哪些人來拉票六”	0	1	0			

## 伍、討論

根據前述實驗結果，以下針對預測力及解析度分別提出數點討論。

### 一、有關預測力的討論

預測力代表平均預測誤差值，從表八可得如下的看法：

1. 基因預測模型對原始的調查資料具有良好的校正能力，不論採用隱藏模式或採用全體模式，其預測力均較原始預測力為佳，第一次民意調查的預測力由2.94%，最佳改善至0.83%，第二次民意調查的預測力由6.40%，最佳改善至1.45%。預測力比值最高達441%，最低亦有192%。
2. 就隱藏模式及全體模式而言，隱藏模式可能達到的預測精度較高，但全體模式的穩定性大，對不同的調查資料均有相類似的預測精度，而且全體模式的平均預測精度（1.49%）略優於隱藏模式（1.70%）。
3. 兩次民意調查的問題大部份相同，不同的部份，解析度也頗低，但兩次民意調查的原始預測力差距極大（ $2.94 : 6.40 = 1 : 2.18$ ），隱藏模式的預測力比值較原始預測力差距更大（ $0.83 : 2.57 = 1 : 3.09$ ），但全體模式能將此差距修正至極小的程度（ $1.53 : 1.45 = 1 : 1.06$ ）。原始預測力的差距大，顯示有外在因素，影響了受訪者對優勢基因組的選擇，隱藏模式承襲了這個影響。但是，雖然有外在因素影響優勢基因組的選擇，受訪者的基本選舉染色體結構並未改變，所以全體模式

表八 預測力對照表

	原始預測力	隱藏模式	全體模式	最佳比	最差比
第一次民意調查	2.94%	0.83%	1.53%	354.22%	192.16%
第二次民意調查	6.40%	2.57%	1.45%	441.38%	249.03%

兩次比值很接近。當外在因素消失，或面臨最後投票抉擇時，受訪者仍然會依其選舉染色體回歸基本投票傾向。因此，隱藏模式似乎較易表現暫時性的變化，而全體模式較能表現長期穩定的傾向。

- 第一次民意調查用隱藏模式的預測效果較佳，第二次民意調查用全體模式的預測效果較佳，可能的改善方法如下。在隱藏模式中，原始表態支持者與候選人間的吸引力未予考慮，也就是假設優勢基因組的影響力大於其他全部基因組的影響力的總和；在全體模式中，優勢基因組則與其他基因組視為相同地位，而由總和決定投票傾向。這兩種觀點或許都太極端，一個可能的改善方式是，將優勢基因組的值乘以一個權數，使其影響力能在總和中適當的提高，但其他基因組的影響力並不完全忽略，若權數的值能適當的決定，則有可能改善整體的預測力。另一個更一般化的改善方式是，每一基因組均給予一權數，使每一基因組的影響力能更正確的表達出來。在本論文中並未考慮權數的決定方法，應可在後續研究中深入探討。
- 基因模型可視為集群分析的一般化模型（generalized model），基因模型與集群分析的主要不同在於：a.基因模型將每一個民意調查的項目均列入考慮，而集群分析僅採用部份的項目，因此集群分析的變數項，較基因模型為少；b.集群分析建立區別方程式作為預測的基礎，而基因模型則建立候選人的染色體作為預測的基礎，兩者在外觀上雖有差異，但其內涵仍可相通，在基因模型中，吸引力的計算實際上是以候選人的基因值為係數，受訪者的基因值為變數，而建立三個（基因模型的）區別方程式，投票的傾向是由（基因模型的）區分方程式的計算值來決定。

## 二、有關解析度的討論

從觀察表七可得出下列分析：

- 依據第三節的分析，優勢基因組解析度的理論值最高為6，表七所示優勢基因組的解析度為6，與第三節所述的最高理論值相符。

2. 兩次民意調查問題解析度的順序大致相符，顯示本論文所採取的解析度分析方法，相當穩定。
3. 第二次民意調查多數問題的回答人數均有增加，顯示愈靠近選舉受訪者的意見愈明確。
4. 三個可能的優勢基因組：明天會投給誰、希望誰當選、比較可能投給誰中，前兩個問題的解析度分列第一、第二，第三個問題則名列五十六。
5. 第一次民意調查中，拉票問題的解析度，幾乎全部列在最後，實不足以作為預測的依據，在第二次民意調查中予以刪除，係一正確處置，以全體模式而言，預測力並未受到影響。
6. 解析度高的前十五個問題，有四個與政黨有關，分列3、11、14、15，其他十一個問題全部與個人形象有關，顯示候選人的人格及政黨形象兩個因素，對受訪者有高度的篩選作用。台北市長選舉視為個人與政黨形象之戰並不為過，執政黨敗選顯示其政黨與候選人的形象有待提昇。
7. 年齡、籍貫、統獨問題、中國人台灣人認同問題緊接在個人與政黨之後名列17至20，這四個問題顯示受訪者主觀的意識型態，而前述候選人形象在統獨及認同問題上又旗幟鮮明，若將候選人的形象解釋為受訪者的意識型態的投射，則統獨及認同問題實相當程度的影響了此次選舉。
8. 三個市政問題分列24、28、31，若扣除拉票問題及戶籍區里問題，市政問題幾乎名列最後。此外，第一個市政問題名列31，顯示不論支持那一候選人，受訪者關心的市政問題是一致的。
9. 性別與使用語言分列41與42，對預測影響不大。
10. 回答戶籍區里問題的人數很少，解析度排名不高，但若考慮平均解析度則排名顯著提昇（參見表九）。

## 陸、結論

本研究依據第一次民意調查的五十八個問題，及第二次民意調查的四十二個問題，分別建立基因預測模型，顯示了相當良好的預測能力。此外本研究尚針對問卷中各個問題分析其解析度，並將各個問題的解析度予以排序，排名較低的問題，對預測的價值也相對較低，若單從預測的角度來看，解析度的分析可作為問題篩選的依據。

基因預測模型的兩種模式（隱藏模式與全體模式）的優劣在本研究中並不明顯，顯示了本研究所提出的預測模型仍然有改善的空間，一個可能的改善方向是對各基因組乘

上權數，加權處理的目的是使各基因組的影響力，能更正確的顯示出來，本研究未能涵蓋這一方面的探討，但可作為後續研究的方向。

表九之一

	問 題	解析度	人數	平均
1	“明天投票會把票投給誰”	6	551	1.0889292
2	“比較希望那一位當選”	5.2487985	666	0.7881079
3	“哪些人來拉票五”	0.0138889	2	0.6944444
4	“對這些問題比較同意那位候選人主張”	2.9859906	515	0.579804
5	“中正區那一里”	0.2158144	38	0.5679327
6	“信義區那一里”	0.3523449	70	0.5033499
7	“這次辯論哪位候選人的表現最好”	2.1868448	443	0.4936444
8	“大安區那一里”	0.4844181	99	0.4893112
9	“最能了解民眾的需要”	3.2124401	674	0.4766232
10	“中山區那一里”	0.2062333	44	0.4687121
11	“哪些人來拉票四”	0.0138889	3	0.462963
12	“松山區那一里”	0.2556803	56	0.4565719
13	“大同區那一里”	0.1277653	28	0.4563048
14	“國民黨、民進黨、新黨較支持哪一黨”	3.8771971	853	0.4545366
15	“做事的能力最強”	3.0703575	680	0.4515232
16	“萬華區那一里”	0.2742849	61	0.4496474
17	“陳水扁、趙少康來比的話，誰會做的比較好”	3.1515296	711	0.4432531
18	“內湖區那一里”	0.2991274	69	0.433518
19	“最有魄力”	3.1047617	723	0.4294276
20	“最清廉不貪汙”	2.4648119	586	0.4206164
21	“文山區那一里”	0.3790086	92	0.4119659
22	“士林區那一里”	0.4253911	106	0.4013123
23	“北投區那一里”	0.379267	98	0.3870072
24	“陳水扁做他會不會做得比黃大洲好”	19.405277	511	0.379751
25	“趙少康做他會不會做得比黃大洲好”	2.2351974	619	0.3610981
26	“南港區那一里”	0.1187418	34	0.3492406
27	“哪些人來拉票三”	0.0416667	12	0.3472222

表九之二

28	“哪些人來拉票二”	0.0951426	28	0.339795
29	“支持不支持民進黨”	2.3309041	783	0.2976889
30	“對黃大洲做市長表現滿意不滿意”	2.4580797	882	0.2786938
31	“支持不支持新黨”	1.986639	720	0.2759221
32	“支持不支持國民黨”	2.180967	920	0.2370616
33	“台北市最需要解決的問題有三”	0.6100844	323	0.1888806
34	“替誰拉票一”	0.3157136	168	0.1879247
35	“民國哪一年出生的”	1.8135583	1023	0.1772784
36	“較偏國民黨、民進黨，還是新黨”	0.6195046	354	0.1750013
37	“籍貫”	1.6995399	1026	0.1656472
38	“哪些人來拉票一”	0.2854085	177	0.1612477
39	“對統獨問題的看法”	1.461402	937	0.1559661
40	“台灣人、中國人或都是”	1.3395847	997	0.1343616
41	“替誰拉票二”	0.0522123	40	0.1305307
42	“台北市最需要解決的問題有二”	0.9206597	706	0.1304051
43	“最高學歷”	1.1375226	1038	0.1095879
44	“投票時是看政黨、候選人還是政見”	0.9861031	941	0.1047931
45	“職業”	1.0767535	1036	0.1039337
46	“替誰拉票三”	0.0079095	9	0.0878829
47	“最喜歡看那一台的電視新聞”	0.4859655	1038	0.0814996
48	“戶籍是在那一區”	0.8426243	1035	0.081413
49	“台北市最需要解決的問題有一”	0.4260764	957	0.0445221
50	“訪問結果”	0.0390996	92	0.0424996
51	“81年立委選舉有沒有投票”	0.4280238	1044	0.0409985
52	“有沒有看台北市市長候選人的電視辯論”	0.2826797	1044	0.0270766
53	“使用的語言”	0.2507355	1044	0.0240168
54	“性別”	0.1125324	1044	0.010779
55	“比較可能投給誰”	0.0093023	100	0.0093023
56	“會不會去投票”	0.0697426	983	0.0070949
57	“有沒有人來拉票”	0.0316537	1042	0.0030378
58	“哪些人來拉票六”	0	1	0

## 參考書目

吳統雄

- 1994 「形象投票模式：選情預測資訊系統應用的實證研究」，**資訊管理**，2卷1期，頁68-81。

洪永泰

- 1994 「選舉預測——一個以整體資料為輔助工具的模型」，**選舉研究**，1卷1期，頁93-110。

陳義彥

- 1989 「台北市選民投票動向的預測初探」，見洪泉湖等編，**發展與省思**，國立清華大學社會科學論文集（一），頁95-121。
- 1994 「我國選民的集群分析及其投票傾向的預測——從民國八十一年立委選舉探討」。 **選舉研究**，1卷1期，頁1-37。

張紘炬

- 1986 「七十五年台北市區域立委選舉得票率統計預測與比較」，**中國統計學報**，24卷12期，頁6-24。

Goldberg, D.E.,

- 1989 *Genetic algorithms in search, optimization and machine learning*, Addison-Wesley Publishing Company, Inc.