

## 應用灰色統計於心臟病資料屬性分析

李嘉陵<sup>1\*</sup>, 蔡坤龍<sup>2</sup>

<sup>1</sup> 銘傳大學生物醫學工程系 中華民國 台灣 桃園

<sup>2</sup> 銘傳大學電子工程學系 中華民國 台灣 桃園

Received 15 April 2010/Revised 15 May 2010/Accepted 16 June 2010

眾所周知與心臟病相關的因素可謂錯綜複雜，如何從中確定重要的因素做為醫師診斷時的參考，是一重要問題。本論文將以灰色統計的方法分析心臟病資料庫中的類別屬性及數值屬性的重要程度，並給予排序。從本論文之分析結果可知在類別屬性方面，心臟病與心絞痛、運動引發心絞痛、心電圖為異常、運動心電圖 ST 間最高斜率有連續兩個以上的 ST 段上升超過 1mm 以上等屬性因素較有關聯，但是空腹血糖值是否大於 120 之屬性與心臟病之關係較小，符合臨床醫學經驗。在數值屬性方面，心臟病與冠狀動脈被阻塞的數量、最大心跳速率與運動心電圖 ST 下降程度等屬性因素較有關聯，亦符合臨床醫學經驗。本論文已證實灰色統計的方法於選擇重要屬性方面是有參考價值的。

關鍵字：灰色統計，白化函數，心臟病症，屬性分析

### 結 論

根據衛生署發表的統計資料顯示，近年來台灣每年大約有 4000 人死於急性心肌梗塞症。若以本病死亡率在 15 ~ 20%，則每年至少有 2 萬 5 千人至 3 萬人罹患心肌梗塞症[1]。由於人口逐年老化以及飲食習慣西化，未來以心血管所衍生的心臟疾病將成為我國重要的醫療問題。雖然有些心臟病，例如先天性心臟病及瓣膜性心臟病，以聽診器就可以明確的判斷出心臟病，但是也有一些必須作胸部 X 光檢查以及心電圖檢查，或甚至其他侵入式的詳細檢查，才能做正確的診斷。但是由於心臟病的相關因素是非常的錯綜複雜，而且經由醫療儀器擷取的數據可能含有「灰色量」，亦即這些數據在某個不確定區間變化，因此如何降低資料中的隨機性，並從中確定重要的因素以便做為醫師診斷時的參考，對現今而言乃是一重要的問題。

為協助醫師探討心臟病檢查項目之重要性，有一些針對大量屬性之分析方法分別被提出：例如以統計分析中的卡方檢定進行類別屬性之相關性檢定，以 T 檢定與 ANOVA 檢定進行數值屬性之分析，或是以資料採礦(Data Mining)技術[8]進行心臟病資料之屬性重要程度分析。本論文擬以灰色系統理論(Grey System Theory)中的灰色統計的方法(Grey Statistics)探討心臟病患者資料集之屬性與心臟病的關聯性，並分別給予類別屬性與數值屬性相對重要程度的排序，做為醫師診斷時的參考。

灰色系統理論是中國華中理工大學鄧聚龍教授於 1982 年所提出。主要是針對系統不明確性、資訊不完整性，亦即所謂「信息不夠及現有數據

不充足」的狀況下，進行關於系統的灰色相關分析與模型建構，並藉著預測和決策等方法來探討及了解系統的特徵和行為。灰色系統理論的主要目的是希望能充分運用諸多灰色系統中的部份已知信息，透過系統分析、灰色生成、灰色建模、灰色預測、灰色統計、灰色決策與灰色控制等方法，來釐清隱藏其中的問題，找尋其中發展規律，進而提出解決之道。灰色預測則是灰色系統理論的重要技術，藉由灰色預測可運用少量的資料去預測未來的資料訊息 [2-5]。灰色預測是以灰色模型(GM, Grey Model)所進行的一種預測方法，其目的是找出某一動態序列數據間的發展關係，而其優點為：(1)建模樣本量小，(2)計算量小、簡便易行，(3)不要求樣本數據需符合常態分配。中國大陸學者陳颯於 1986 年提出了灰色預測控制器[2, 3]來控制未知的系統之後，灰色預測的應用便開始快速的發展。在 2002 年 Kuen-Yih Shy 等人[6]所發表之論文是利用灰色預測的方法來搭配自我調整控制器，因為自我調整控制器必需藉由輸出入的資料來設計，所以利用灰預測來預測輸出入的資料，Shy 等人的研究結果發現利用灰預測可產生自我調整時所需的資料，也能夠濾除雜訊使系統更穩定。在 2001 年 Chien-Hsin Chou[7]也發表一篇利用灰預測去對實際系統與目標系統之間的誤差做事先的調整，使得系統更穩定。

本論文之章節安排如下：下一節將本論文擬解決之問題做一描述，包含心臟病患者資料集之介紹與本論文之主要目標。第 3 節將說明類別屬性與數值屬性的灰色統計分析法。結果與討論以及最後之結論分別於第 4 節與第 5 節說明。

### 材料及方法

#### 問題描述

本論文以葡萄牙波多大學(University of Porto)人工智慧及電腦科學實驗室所分享的心臟病患者資料集[9]為研究對象，提出灰色統計方法將此資料

\*通訊作者連絡地址：銘傳大學生物醫學工程系，333 桃園縣龜山鄉德明路 5 號，電話：(03)350-7001 ext 3549，傳真：(03)359-3840，e-mail: [jllee@mail.mcu.edu.tw](mailto:jllee@mail.mcu.edu.tw)

集的類別與數值屬性共 13 個條件屬性之相對重要程度排序，做為醫師診斷之參考。此資料集具有 1 個目標屬性與 13 個條件屬性，包含 7 個類別屬性及 6 個數值屬性，共有 270 筆對象資料。該資料集可區分為患有心臟病(120 位)及無心臟病(150 位)等兩類資料集，以下將這些條件屬性依類別屬性及數值屬性分別說明於表一與表二，其中屬性序號為原資料集之序號。為便於與其他研究方法比較，本論文採用文獻 [8] 的做法，將 270 位病患之原始資料以分層(無/有心臟病)抽樣方式進行如下：

1. 將 150 位無心臟病患者之資料隨機分成兩群 group A: 80%(120 筆資料)及 group B: 20%(30 筆資料)。
2. 將 120 位有心臟病患者之資料亦隨機分成兩群 group C: 80%(96 筆資料)及 group D: 20%(24 筆資料)。
3. 將 group A 及 group C 的資料組合成為訓練資料：120 筆無心臟病與 96 筆有心臟病共 216 筆資料。
4. 將 group B 及 group D 的資料組合成為測試資料：30 筆無心臟病與 24 筆有心臟病共 54 筆資料。

上述之分法已確保訓練與測試資料中無/有心臟病之比例(150:120=55.56%:44.44%)與母體相同。本論文將 group A 及 group C 組合成訓練資料共 216 筆，依據類別屬性及數值屬性分別將灰色統計理論適當的應用於心臟病資料庫中的屬性相對重要程度分析中，主要關鍵是應用灰色統計之白化函數(Whitening function)將原始資料映射至灰矇矇集以便進一步計算灰統計值以及屬性相對重要程度值，並據此給予排序。

### 灰色統計之屬性分析法

進行資料屬性分析前，必須先確認資料品質之良窳及重要屬性之完整性，因為若資料有空值(Missing values)或錯誤值等，則會造成分析結果不正確[10]。本論文所取得之資料集有兩筆資料的屬性很容易被誤認，其一是歸類為有病的資料列向量中較具重要性的三種屬性：血管阻塞數、胸痛、運動時引發心絞痛之屬性值皆為 0，只有心電圖 ST 段異常下降屬性值呈現偏高的情形；其二是無病的資料列向量，其血管阻塞數卻有極度危險的 3 條，心電圖 ST 段異常下降也是屬於危險的 1.8。但謝醫師認為心臟病患者有其特殊性[11]，因此仍不能將此兩筆資料去除。灰色統計之分析法是對各個統計對象有關項目的樣本進行處理以瞭解項目所屬灰類[2]：設 I 為評估對象指標集  $I=\{1, 2, \dots, w\}$ ，J 為評估項目指標集  $J=\{1, 2, \dots, m\}$ ，K 為灰類指標集  $K=\{1, 2, \dots, n\}$ ，又設  $d_{ij}$ ， $i \in I$ ， $j \in J$  為對象 i 對於項目 j 的樣本值，則可得樣本矩陣 d 為  $d=[d_{ij}]$ ， $i \in I$ ， $j \in J$ ，以及設  $f_k$  為 k 灰類之白化函數(Whitening function)。本論文採用之 k 灰類白化函數包含下列三種：(1) "1 灰類"白化函數如圖一所示，其中(d0, d20)為折點值，亦即  $d_{ij} > d20$ ，則"1 灰類"之默認度定義為 1，(2) "2 灰類"白化函數如圖二所示，其中(d10, d30)為折點值，亦即  $d_{ij} < d10$ ，則"2 灰

類"之默認度定義為 1，(3) "3 灰類"白化函數如圖三所示，其中(d10, d51)為折點值，亦即  $d_{ij} < d10$  及  $d_{ij} > d51$ ，則"3 灰類"之默認度定義為 1。本論文將樣本數據運用上述白化函數映射以便進一步計算出灰統計值與屬性相對重要程度值，以便給予排序。因此評估項目所屬灰類是依臨床醫學之知識先行決定如下：由於類別屬性是以統計方法計算，在有心臟病與無心臟病之類別中皆是以其統計量表示其支持度，因此評估項目採用"1 灰類"之白化函數。然而，數值屬性須依其特性而用不同之白化函數，採用之原則是讓某項目若屬於健康範圍，則其值為小，如表三所示。於下列兩節中將具體詳述如何計算類別屬性及數值屬性之屬性相對重要程度值。

**類別屬性分析法** 本節針對訓練用的心臟病患者資料集的類別屬性 chest 說明灰統計屬性分析法，本方法除灰色默認度之映射外，其餘步驟參照[8]之做法。屬性 chest 有 4 個屬性值(A31, A32, A33, A34)，屬性描述如表一所示，灰統計屬性分析法說明如下：

頻率統計：頻率表示某一個屬性值在各自類別中出現次數。

支持度之計算：支持度表示某一個屬性值在各自類別中出現之機率。

默認度之映射：將各自支持度經由"1 灰類"之白化函數映射。

信賴度之計算：在各自類別中的默認度除以兩類別默認度總和。

屬性相對重要程度之計算：每一屬性值之相對重要程度依下列公式計算： $ABS(\text{無病信賴度}-\text{有病信賴度}) * ((\text{無病頻率} + \text{有病頻率}) / \text{總人數})$ ，再將該屬性所屬之每一屬性值之相對重要程度相加，其中 ABS 為取絕對值函數。此式之含意為信賴度差異愈大，頻率愈高，則屬性相對重要程度愈高。

類別屬性 chest 分析的結果如表四，其餘 6 個類別屬性依此類推，其結果如表五。舉例說明表四之屬性相對重要程度之計算過程：屬性值 A31 在無心臟病類別中出現 13 次，其支持度為  $0.108=13/120$ ，在有心臟病類別中出現 4 次，其支持度為  $0.042=4/96$ ，無心臟病之默認度之映射為  $0.128=0.108/0.844$ ，有心臟病之默認度之映射為  $0.049=0.042/0.844$ ，無心臟病之信賴度為  $0.723=0.128/(0.128+0.049)$ ，有心臟病之信賴度為  $0.276=0.049/(0.128+0.049)$ ，最後 A31 之屬性相對重要程度因而為  $0.035=(0.723-0.276)*(13+4)/(120+96)$ 。

以上默認度之折點值皆取 0.844，是因為有心臟病(C group)類別中所有類別屬性的支持度(即機率值)的最大值為 0.844，以此值為默認度之折點值，使得默認度之範圍仍在 [0, 1] 之間。

**數值屬性分析法** 本節針對訓練用的心臟病患者資料集的數值屬性說明灰統計的數值屬性分析法，其具體分析法如下步驟：

1. 決定白化函數及折點值：在各自類別中樣本矩陣為  $d=[d_{ij}]$ ， $i \in I$ ， $j \in J$ ，決定每一屬性之灰類白化函數及折點值，如表三所示。

2. 默認度之映射：將樣本矩陣代入白化函數計算每一屬性值之默認度。
3. 默認度之平均值計算：由於兩類樣本矩陣之人數並不一致，因此以每一屬性值之默認度平均值代表該屬性值。
4. 灰統計值 (delta) 計算：計算公式為  $\text{delta}=\sqrt{\text{abs}(f\_avg(1)-f\_avg(2))}$ ，其中  $f\_avg(i)$  為兩類各自默認度之平均值，其中  $i=1, 2$  表示無心臟病與有心臟病的兩類別， $\text{abs}()$  為取絕對值函數， $\text{sqrt}()$  為開根號函數。以此灰統計值代表數值屬性之屬性相對重要程度。

結果如圖四，其餘 5 個數值屬性依此類推，其結果如表七之 GS 方法。

屬性序號	屬性名稱 (代號)	屬性值	屬性描述
屬性 2	sex	A21	病患性別 A21 女性 A22 男性
		A22	
屬性 3	chest	A31	胸痛類型 A31 主體脈刺激引起主動脈瘤急性發作時，所引發的疼痛 A32 心包膜炎引起的胸痛，深呼吸時會使疼痛加重 A33 官能性神經症所引發的疼痛 A34 心絞痛
		A31	
		A33	
		A34	
屬性 6	fasting_blood_sugar (FBS)	A61	空腹血糖值是否大於 120 A61 否 A62 是
		A62	
屬性 7	resting_electro_cardiography (ECG)	A71	休息時心電圖 A71 正常 A72 輕微異常 A73 異常
		A72	
		A73	
屬性 9	exercise_induced_angina (EIA)	A91	運動是否引發心絞痛 A91 否 A92 是
		A92	
屬性 11	slope	A111	運動心電圖 ST 間最高斜率 A111 完全正常 A112 連續兩個以上的 ST 段上升超過 1mm 以上 A113 連續兩個以上的 ST 段下降超過 1mm 以上
		A112	
		A113	
屬性 13	thal	A131	地中海型貧血 A131 正常 A132 地中海型貧血甲型 A133 地中海型貧血乙型
		A132	
		A133	

表一、類別屬性描述

屬性序號	屬性名稱 (代號)	屬性描述
屬性 1	age (age)	年齡
屬性 4	resting_blood_pressure (BP)	休息時血壓
屬性 5	serum_cholesterol (SC)	血液中膽固醇含量
屬性 8	maximum_heart_rate_achieved (MHR)	最大心跳速率
屬性 10	oldpeak (OP)	運動心電圖 ST 下降程度
屬性 12	number_of_major_vessels (NMV)	冠狀動脈被阻塞的數量

表二、數值屬性描述

屬性序號	屬性名稱 (代號)	灰類白化函數	折點值
屬性 1	age (age)	"1 灰類"	(0, 77)
屬性 4	resting_blood_pressure (BP)	"3 灰類"	(80, 140)
屬性 5	serum_cholesterol (SC)	"1 灰類"	(50, 220)
屬性 8	maximum_heart_rate_achieved (MHR)	"2 灰類"	(100, 202)
屬性 10	oldpeak (OP)	"1 灰類"	(0, 6.2)
屬性 12	number_of_major_vessels (NMV)	"1 灰類"	(0, 3)

表三、數值屬性所屬灰類白化函數

屬性名稱	屬性值	無心臟病(A group)				有心臟病(C group)				屬性相對重要程度
		頻率	支持度	默認度	信賴度	頻率	支持度	默認度	信賴度	
chest	A31	13	0.108	0.128	0.723	4	0.042	0.049	0.276	0.035
	A32	29	0.242	0.286	0.885	3	0.031	0.037	0.115	0.114
	A33	43	0.358	0.425	0.683	16	0.167	0.197	0.317	0.100
	A34	35	0.292	0.346	0.277	73	0.760	0.901	0.723	0.223
SUM		120		1.185	2.567	96		1.185	1.433	0.472

表四、chest 類別屬性分析結果

屬性名稱	屬性重要程度 1 (折點值 0.606)	排序 1	屬性重要程度 2 (折點值 0.875)	排序 2
thal	0.405	2	0.478	1
chest	0.424	1	0.472	2
slope	0.38	3	0.420	3
EIA	0.29	4	0.409	4
sex	0.21	5	0.313	5
ECG	0.199	6	0.215	6
FBS	0.007	7	0.015	7

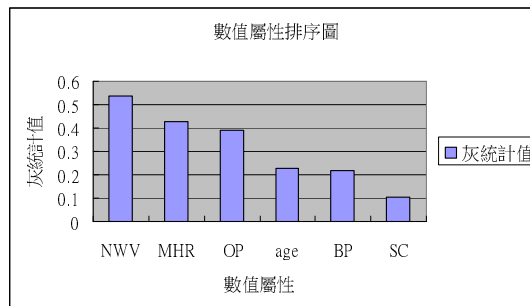
表五、類別屬性重要程度比較表

屬性名稱	GS方法 屬性相對重要程度 (折點值 0.875)	GS方法 排序	DM方法 屬性相對重要程度	DM方法 排序	卡方檢定法 卡方值	卡方檢定法 P-value	卡方檢定法 排序
thal	0.478	1	0.505	1	56.897	4.415E-13	1
chest	0.472	2	0.493	2	51.855	3.215E-11	2
slope	0.420	3	0.327	4	23.324	8.615E-06	4
EIA	0.409	4	0.334	3	29.235	6.410E-08	3
sex	0.313	5	0.301	5	21.814	3.004E-06	5
ECG	0.215	6	0.207	6	9.151	0.010	6
FBS	0.015	7	0.006	7	0.017	0.895	7

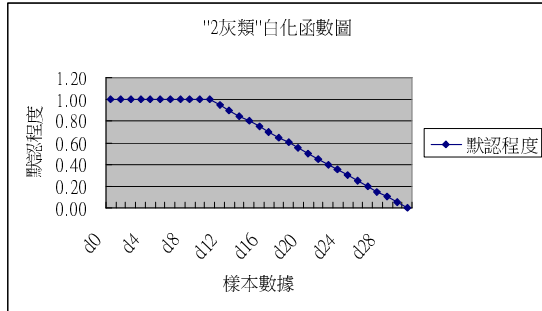
表六、3 種方法之類別屬性相對重要程度分析比較表

屬性名稱	GS方法 屬性相對重要程度	GS方法 排序	DM方法 屬性相對重要程度	DM方法 排序	T檢定法 T-value	T檢定法 P-value	T檢定法 排序
NMV	0.538	1	0.969	1	6.776	0.000	1
MHR	0.427	2	0.912	2	6.586	0.000	2
OP	0.390	3	0.836	3	5.813	0.000	3
age	0.227	4	0.326	4	2.391	0.018	4
BP	0.221	5	0.227	5	0.232	0.817	6
SC	0.104	6	0.055	6	0.392	0.696	5

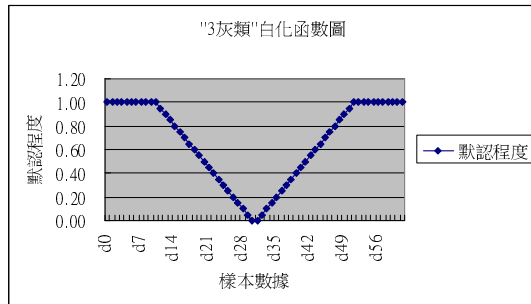
表七、3 種方法之數值屬性相對重要程度分析比較表



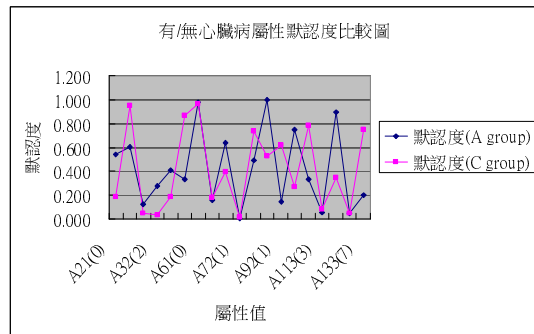
圖一、"1 灰類"白化函數圖



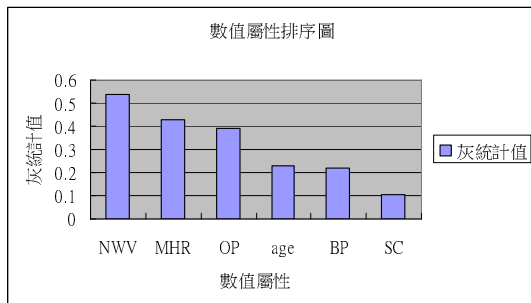
圖二、“2 灰類”白化函數圖



圖三、“3 灰類”白化函數圖



圖四、有心臟病與無心臟病類別屬性默認度比較圖



圖五、數值屬性相對重要程度排序圖（屬性代號如表二）

**結果與討論**

由圖四的屬性默認度比較圖可以得知以下訊

息：

女性在無心臟病之屬性默認度及男性在有心臟病之屬性默認度皆較高，表示男性得心臟病之機率較女性高。

心絞痛（A34）在有心臟病之病患其屬性默認度偏高。亦即，當胸痛類型是因心絞痛而引起，則被歸類為有心臟病之可能性很高。

空腹血糖值是否大於 120 與心臟病之關係較小。心電圖正常（A71）而有心臟病者較少，大多數心臟病者之心電圖為異常（A73）而且運動會引發心絞痛（A92）以及運動心電圖 ST 間最高斜率有連續兩個以上的 ST 段上升超過 1 mm 以上（A112）。

心臟病者得地中海型貧血乙型（A133）者亦較多，這可能是葡萄牙地區性的情形。

從表五類別屬性相對重要程度排序比較可以得知：白化函數(Whitening function)之折點值愈大，則屬性相對重要程度值亦愈大；白化函數之折點值對於屬性相對重要程度值之排序有所影響，但重要屬性仍然歸屬於重要。於圖五中可知數值屬性之屬性相對重要程度排序依次為 1. NWV（冠狀動脈被阻塞的數量） 2. MHR（最大心跳速率） 3. OP（運動心電圖 ST 下降程度） 4. age（年齡） 5. BP（休息時血壓） 6. SC（血液中膽固醇含量），由此可知，心臟病與冠狀動脈被阻塞的數量、最大能達到之心跳速率與運動心電圖 ST 下降程度等因素最有關連，符合臨床醫學經驗。

在類別屬性重要程度分析法中，傳統上有以卡方檢定進行相關性檢定。在數值屬性重要程度分析法中，傳統上有以 T 檢定與 ANOVA 檢定進行重要性分析。近年來有以資料採礦 (Data Mining, DM) 進行類別屬性與數值屬性相對重要程度分析（以下簡稱為 DM 方法 [8]）。本論文提出灰統計 (Grey Statistics, GS) 分析類別屬性與數值屬性相對重要程度（以下簡稱為 GS 方法）。以下將比較這些方法之差異，其結果如表六與表七所示。表六為 3 種方法之類別屬性相對重要程度分析比較。以卡方檢定法計算各個類別之條件屬性與目標屬性間的相關程度，須先計算其卡方值，之後才以此值查表算出 P-value。P-value 愈小，表示此屬性愈重要。在 95% 信心水準之下，P-value < 0.05 則為重要之屬性。因此從表六可知卡方檢定法僅過濾掉 FBS 屬性（空腹血糖值），亦即選擇前 6 個屬性為重要屬性；在此 6 個重要屬性中，DM 方法及本論文所提出 GS 方法皆位居前六名，表示本文 GS 方法的有效性。在排序上本論文所提出 GS 方法在 slope 屬性略為不同，然而在醫學上已證實 ST 段上升超過 1 mm 以上已表示心肌梗塞，ST 段下降超過 1 mm 以上已表示冠狀動脈痙攣現象 [11]，其嚴重程度不亞於運動引發心絞痛，因此本論文所提出 GS 方法於選擇重要屬性方面頗有參考價值。表七為 3 種方法之數值屬性相對重要程度分析比較。T 檢定法之 P-value 愈小，表示此屬性愈重要。在 95% 信心水準之下，P-value < 0.05 則為重要之屬性。因此從表七可知 T 檢定法會過濾掉 BP 屬性（休

息時血壓)與 SC 屬性(膽固醇含量),亦即選擇前 4 個屬性為重要屬性。在此 4 個重要屬性中,DM 方法及本論文所提出 GS 方法皆位居前 4 名,表示本文 GS 方法的有效性。

### 結 論

本論文以灰色統計的方法將心臟病資料庫的資料依各屬性在其有病及無病兩類別的差異性而給予排序。從以上之分析可知,不論類別屬性或數值屬性,本論文的方法皆能有效地篩選出重要屬性。在類別屬性方面,心絞痛、運動引發心絞痛、心電圖異常以及運動心電圖 ST 間最高斜率有連續兩個以上的 ST 段上升超過 1 mm 以上等屬性因素較能突顯兩類別的差異,但是空腹血糖值是否大於 120 之屬性的差異性較小,符合臨床醫學經驗。在數值屬性方面,心臟病與冠狀動脈被阻塞的數量、最大之心跳速率與運動心電圖 ST 下降程度等屬性因素較有關聯,亦符合臨床醫學經驗,由上可知本方法的正確性。經由本論文灰統計白化函數之折點值試驗,我們推論對於不同樣本,屬性重要程度值將會有所不同,但針對心臟病之屬性重要程度之排序約略一致。因此本論文所提出灰色統計方法於選擇重要屬性方面有其參考價值。

### 參考文獻

1. 何逸德 2002, 病理檢驗醫學, 力大圖書, 台北, 第 91-95 頁。
2. 鄧聚龍 2003, 灰色系統理論與應用, 高立圖書, 台北, 第 1-117 頁。
3. 翁慶昌, 陳嘉權, 賴宏仁 2001, 灰色系統基本方法及其應用, 高立圖書, 台北, 第 114-118 頁。
4. 施昆毅, 吳松霖, 王銀添 2001, Internal Model Control with a Grey Predictor, 灰色系統理論與應用學術研討會, 雲林, 第 113-119 頁。
5. 黃景祺 2000, 灰差分預測器之設計及應用, 碩士論文, 國立台灣師範大學工業教育研究所, 台北。
6. Chou, C. H. 2001, A Variable Structure Controller Based on The Grey Prediction Technology, Proceedings of the American Control Conference, Arlington, p.1505-1506.
7. Shy, K. Y., and Wang, Y. T. 2002, Grey Modeling Control of a Thermal Barrel in Plastic Molding Process, Proceedings of the World Congress on Intelligent Control and Automation, Shanghai, p.2557-2560.
8. Lee, Y. S., et al 2007, A Data Mining Approach to Constructing Heart Disease Forecast Model, Journal of Informatics and Electronics (JIE), 2 :19-28.
9. <http://www.liacc.up.pt/>
10. Lee, Y. S., et al 2006, Data Mining on Credit Risk of Basel Capital Accord II: The Construction of Collection Score Card, Innovation and Management.
11. 謝孟亨 2009, 心臟病資料集之面談, 長庚醫院呼吸胸腔內科主治醫師。

## Application of Grey Statistics to Data Attribute Analysis From Heart Disease Set

Jia-Ling Lee<sup>1\*</sup>, Steven Tsai<sup>2</sup>

<sup>1</sup>*Department of Biomedical Engineering, Ming Chuan University, Taiwan, R.O.C.*

<sup>2</sup>*Department of Electronic Engineering, Ming Chuan University, Taiwan, R.O.C.*

Received 15 April 2010/Revised 26 May 2010/Accepted 13 June 2010

It is well known that the factors related to the heart disease are complex. How to identify some important factors to act as the diagnostic considerations for doctors is an important problem. This paper will propose a novel method of grey statistics to analyze data attributes in a heart disease database, including class attributes and numerical attributes. The purpose of this paper is then to distinguish some important data attributes related to the heart disease and to rank these attributes. We find that the heart disease is more relation to some class attributes including angina pectoris, exercise\_induced angina, abnormal electrocardiography and abnormal slope of ST segment in ECG, and some numerical attributes including the number of major vessels oppilated, maximum heart rate achieved and the drop level in the slope of ST segment in ECG. These results of this paper are all coincidence with the clinical knowledge. We have verified that the method of grey statistics is valuable to sort some more important attributes.

Keywords: Grey statistics, Whitening function, Heart disease, Attribute analysis.