

# 國立交通大學

電機學院 電機與控制學程

## 碩士論文

圖形識別與類神經網路  
於 2006 年世界盃足球賽的分類及預測

Pattern Recognition and Neural Networks for  
Classification and Prediction of 2006 World Cup  
Football Game



研究生：張文龍

指導教授：林進燈 教授

黃國源 教授

中華民國九十六年一月

圖形識別與類神經網路於2006年世界盃足球賽的分類及預測  
Pattern Recognition and Neural Networks for  
Classification and Prediction of 2006 World Cup Football Game

研究生：張文龍

Student: Wen-Lung Chang

指導教授：林進燈博士

Advisor: Dr. Chin-Teng Lin

黃國源博士

Dr. Kou-Yuan Huang

國立交通大學  
電機學院 電機與控制學程  
碩士論文



Submitted to College of Electrical and Computer Engineering

National Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of

Master of Science

in

Electrical and Control Engineering

January 2007

Hsinchu, Taiwan

中華民國九十六年一月

# 圖形識別與類神經網路於2006年世界盃足球賽的分類及預測

學生：張文龍

指導教授：林進燈 教授  
黃國源 教授

國立交通大學 電機學院 電機與控制學程碩士班

## 摘 要

本論文主要研究的方向有兩個，第一個研究方向為運用圖形識別中的各種非監督式分類法來分析 2006 年世界盃足球賽的球隊實力分為幾類，分類方法有 K-Means 分類演算法、Fuzzy C-Means 分類演算法、Hierarchical 分類演算法、及 Self-Organizing Feature Map 分類演算法，經由 2006 年世界盃足球賽官方網站所得到的統計數據，除分析各種非監督式分類法的分類正確率及分類結果之外，同時也針對數據資料作分類個數的有效性驗證。第二個研究方向為運用監督式類神經網路中的多層感知器及倒傳遞學習演算法，並根據 2006 年世界盃足球賽前一階段的比賽數據來預測下一階段比賽兩隊的勝率，經由實驗結果，本論文建構的勝率預測模型準確率達 62.5%，若不含比賽結果平手的場次，則準確率達 76.9%。

# Pattern Recognition and Neural Networks for Classification and Prediction of 2006 World Cup Football Game

Student : Wen-Lung Chang

Advisors : Dr. Chin-Teng Lin  
Dr. Kou-Huang Huang

Degree Program of Electrical and Computer Engineering  
National Chiao Tung University

## ABSTRACT

There are two main research aspects in this thesis. The first research aspect is that we use 4 unsupervised clustering methods to analyze the team level of 2006 FIFA World Cup Football Games. These clustering methods include K-means clustering algorithm, Fuzzy C-means clustering algorithm, hierarchical clustering algorithm, and self-organizing feature map algorithm. Furthermore, we use 3 clustering validity methods to verify whether these football statistical data possesses a clustering property. The second research aspect is that we use the supervised multi-layer perceptron with back propagation learning rule to predict the winning rate based on two team's previous football game's records. According to the experimental results, it shows that the correct rate achieves 62.5 % by using the adopted prediction model. If the draw games are excluded, the correct rate can achieve 76.9 %.

## 誌 謝

首先我要感謝我的指導教授，林進燈博士，在研究方向上給我很大的自由度，再來特別要感謝我的良師益友指導教授，黃國源博士，這二年來的指導，尤其是在最後幾個月論文趕工的日子裏，不時地給予我意見及實驗上的指導，讓我終於能順利地完成我的碩士學位論文，這過程中也讓我更深深地體會黃教授在學術上嚴謹的研究態度。

再來我要感謝實驗室的學弟：陳楷儒、董安晉、周映良、謝岳勳、及陳俊宇，謝謝你們在實驗室的這段期間內，給予我適時的解惑及課業上必要的協助。

還有我要謝謝我公司的長官們，容許及配合我留職停薪半年，得以專心地重拾書本完成學業，讓我終於可以放下五年多以來在我心頭的一顆大石頭(是壓力也是心願)，真是如釋重負。

最後我要感謝我的媽媽在家的辛勞，我的弟弟在經濟上的支援，我的學長好朋友 Twokey 在經驗上的建議、及 Andrew 在精神上的鼓勵。

謝謝你們！

中文摘要	.....	i
英文摘要	.....	ii
誌謝	.....	iii
目錄	.....	iv
表目錄	.....	vi
圖目錄	.....	vii
第一章	圖形識別與類神經網路於 2006 年世界盃足球賽球隊實力之分類	1
1	介紹	1
2	採用的非監督式分類法簡介	4
2.1	K-Means 分類法	4
2.2	Fuzzy C-Means 分類法	5
2.3	凝聚階層式分類法	6
2.4	Self-Organizing Feature Map Neural Networks(SOFM)分類法	7
2.5	順序性群聚分類演算法	8
3	採用的分類有效性驗證法	12
3.1	以 BSC 及 MBSC 為基礎的有效性驗證法	13
3.2	以 K-Means 為基礎的有效性驗證法	14
4	非監督式分類法用於 2006 年世界盃足球賽球隊實力的分類	16
4.1	原始資料的擷取及轉換	16
4.2	比賽結果的分類原則	21
4.3	分類正確率計算	22
5	實驗結果	24
5.1	非監督式分類法的實驗結果	24
5.2	2006 年世界盃足球賽統計數據分類個數有有效性的驗證結果	34
5.3	主分量分析法用於 2006 年世界盃足球賽統計數據之分類實驗結果	36
6	結論	44
7	參考文獻	45
第二章	多層感知器類神經網路於 2006 年世界盃足球賽的勝率預測	46
1	介紹	46
2	多層感知器(MLP)及倒傳遞(BP)學習演算法	50
2.1	多層感知器(MLP)網路架構	50
2.2	倒傳遞(BP)學習規則	50
3	2006 年世界盃足球賽的勝率預測	53
3.1	原始資料的擷取與轉換	53
3.2	網路訓練的目標值	54
3.3	訓練資料的選取及預測隊伍之輸入資料的轉換	55
3.4	MLP 的 BP 學習法則參數的決定	57
4	2006 年世界盃足球賽勝率預測的結果與討論	61
4.1	勝率預測結果	61
4.2	預測結果的討論	64
5	結論	67

6	參考文獻 .....	68
附錄一	2006 世界盃足球賽賽程表 .....	69
附錄二	2006 世界盃足球賽單場比賽報告 .....	71
附錄三	2006 世界盃足球賽比賽結果 .....	72





## 表 目 錄

第一章	圖形識別與類神經網路於 2006 年世界盃足球賽球隊實力之分類 . . . . .	1
表 4-1	德國隊(GER)對哥斯大黎加隊(CRC) 17 項原始資料 . . . . .	17
表 4-2	德國隊(GER)對哥斯大黎加隊(CRC) 8 項原始資料的擷取與正規化轉換 . .	17
表 4-3	32 隊全部比賽的原始資料 . . . . .	18
表 4-4	32 隊原始資料經轉換後的資料 . . . . .	19
表 4-5	資料轉換後 32 隊,把 32 隊各自資料加起來後的資料 . . . . .	20
表 4-6	32 隊各隊的單場平均輸入資料 . . . . .	20
表 4-7	2006 年世界盃足球賽「比賽結果」分類對照表 . . . . .	21
表 4-8	分類結果比對表 . . . . .	22
表 4-9	K-Means 演算法分類結果表 . . . . .	23
表 4-10	K-Means 演算法分類結果比對表 . . . . .	23
表 5-1	K-Means 分類結果表 . . . . .	25
表 5-2	K-Means 分類結果的各類中心 . . . . .	25
表 5-3	K-Means 分類的正確隊數及正確率 . . . . .	25
表 5-4	合併式階層分類結果表 . . . . .	28
表 5-5	合併式階層分類的正確隊數及正確率 . . . . .	28
表 5-6	Fuzzy C-Means 分類結果表 . . . . .	29
表 5-7	Fuzzy C-Means 分類的正確隊數及正確率 . . . . .	30
表 5-8	SOFM 分類結果表 . . . . .	32
表 5-9	SOFM 分類正確隊數及正確率 . . . . .	32
表 5-10	非監督式群聚分類法實驗結果之正確率比較表 . . . . .	33
表 5-11	2006 世界盃足球賽分類輸入資料之 principal component and feature coefficient . . . . .	36
表 5-12	2006 世界盃足球賽分類輸入資料之 covariance matrix 的 eigenvalues 及 total eigenvalues 百分比,及 total eigenvalues 累積的百分比 . . . . .	36
表 5-13	PCA 處理後取不同 feature 個數, K-Means 及 Fuzzy C-Means 分類結果的正確隊數比較表 . . . . .	38
表 5-14	未經 PCA 分析之 2006 世界盃足球賽的統計資料分成 3 類的分類結果 . . . . .	40
表 5-15	取 4 個 features 經 PCA 分析後之 2006 世界盃足球賽的統計資料分成 3 類的分類結果 . . . . .	41
表 5-16	未經 PCA 分析之 2006 世界盃足球賽的統計資料分成 2 類的分類結果 . . . . .	42
表 5-17	取 4 個 features 經 PCA 分析後之 2006 世界盃足球賽的統計資料分成 2 類的分類結 . . . . .	43
第二章	多層感知器類神經網路於 2006 年世界盃足球賽的勝率預測 . . . . .	46
表 1	德國隊(GER)對哥斯大黎加隊(CRC) 17 項原始資料 . . . . .	54
表 2	德國隊(GER)對哥斯大黎加隊(CRC) 8 項原始資料的擷取與正規化轉換 . .	54
表 3	預測德國隊(GER)對瑞典隊(SWE)的輸入資料 . . . . .	56
表 4	決定三種參數使用值之測試條件 . . . . .	57
表 5	第二階段到第五階段的預測結果表 . . . . .	63
	第二階段到第五階段不含平手場次的預測正確率 . . . . .	66



## 圖 目 錄

第一章	圖形識別與類神經網路於 2006 年世界盃足球賽球隊實力之分類 . . . . .	1
圖 1-1	非監督式圖形識別於 2006 年世界盃足球賽球隊的分類及分類有效性的系統 . . . . .	2
圖 1-2	2006 年世界盃足球賽 64 場賽程表 . . . . .	2
圖 1-3	所採用的非監督式分類法 . . . . .	3
圖 1-4	所採用的分類有效性驗證法 . . . . .	3
圖 2-1	SOM 網路架構 . . . . .	8
圖 2-2	3 種常用的輸出神經元的 Topology . . . . .	8
圖 2-3	同一樣本，在不同時間出現，經 BSC 分類的結果 . . . . .	10
圖 4-1	分類示意圖 . . . . .	21
圖 5-1	10,000 次 K-Means 分類的正確隊數統計柱狀圖 . . . . .	25
圖 5-2	合併式階層分類結果樹狀圖 . . . . .	26
圖 5-3	10,000 次 Fuzzy C-Means 分類的正確隊數統計柱狀圖 . . . . .	29
圖 5-4	2006 年世界盃足球賽 SOFM 分類演算法分成 4 類的 output nodes topology . . . . .	31
圖 5-5	$\eta_j^*(t)$ 與 t 的關係圖 . . . . .	31
圖 5-6	200 次 SOFM 正確隊數統計柱狀圖 . . . . .	31
圖 5-7	BSC 分類有效性結果圖 . . . . .	35
圖 5-8	MBSC 分類有效性結果圖 . . . . .	35
圖 5-9	PFS 分類有效性結果圖 . . . . .	35
圖 5-10	2006 年世界盃足球賽輸入資料經 PCA 分析後投影至 8 個主軸上，total eigenvalues 所累積的百分比直條圖 . . . . .	37
圖 5-11	經 PCA 分析後取不同 feature 個數，K-Means 及 Fuzzy C-Means 分類結果的正確隊數曲線圖 . . . . .	38
第二章	多層感知器類神經網路於 2006 年世界盃足球賽的勝率預測 . . . . .	46
圖 1	Supervised Prediction System . . . . .	47
圖 2	2006 年世界盃足球賽的訓練資料及預測的階段之關係圖 . . . . .	48
圖 3	2006 年世界盃足球賽第一階段績分圖 . . . . .	49
圖 4	預測 2006 年世界盃足球賽勝率所採用的 MLP 網路架構 . . . . .	50
圖 5	2 ~ 40 Hidden neurons 與 3 種 MSE 的 epochs 比較圖 . . . . .	58
圖 6	設定 MSE=0.01， $\eta=0.1$ ，hidden neurons=11，測試 3 種 $\beta$ 與 epochs 的比較圖 . . . . .	59
圖 7	設定 MSE=0.01， $\beta=0.6$ ，hidden neurons=11，測試 5 種 $\eta$ 與 epochs 的比較圖 . . . . .	60
圖 8	MLP 的 BP 的訓練過程圖 . . . . .	61

# 第一章 圖形識別與類神經網路於 2006 年世界盃足球賽球

## 隊實力之分類

### 摘要

群聚分類理論在資料分析中佔有相當重要的角色，我們利用 2006 年世界盃足球賽官方統計數據的各項技術指標及圖形識別中的非監督式分類法來將球隊作分類，所用的方法有 K-Means 分類演算法、Fuzzy C-Means 分類演算法、Hierarchical 分類演算法、及 Self-Organizing Feature Map 類神經網路，根據競賽規則的特性把球隊分成 4 類，分類結果顯示各種分類法於 2006 年世界盃足球賽的分類正確率約介於 46.87 % ~ 68.75 %。我們再利用 Basic Sequential Clustering, Modified Basic Sequential Clustering 為基礎的有效性驗證法，及以 K-Means 為基礎的 Pseudo F-Statistic 公式，共 3 種分類有效性驗證法，針對 2006 年世界盃足球賽統計數據的最佳分類個數進行有效性驗證，驗證結果顯示 2006 年世界盃足球賽官方數據以分成 2 類或 3 類為較佳之分類個數，最後我們將 2006 年世界盃足球賽資料經過主分量分析 (PCA) 後，把原始資料投影到前四個主分量軸上後，再重新分成 2 類及 3 類的情況下，明顯地看出強隊與弱隊可被區分開。

### 1. 介紹

2006 年世界盃足球賽會內賽開幕前，黃國源、張生平、陳楷儒、董安晉、...等等，曾根據會前賽的統計數據提出以類神經網路的方法分析 2006 年世界盃足球賽球隊的實力 [1]，而 2006 年德國世界盃會內賽結束後，從官方網站 [2] 中，我們可獲得由世界足球聯盟組織(FIFA) [3] 針對球隊的攻守數據之統計而得到各項足球技術的統計資料，例如：球隊進球數、失球數、控球時間、十二碼罰球數、紅牌數...等等，共 17 項，每一項統計資料皆代表一項技術指標，而這些技術指標對球賽最終結果都具有影響力，因此針對這些統計資料，我們利用圖形識別與類神經網路中的非監督式分類法，將資料分為數個類別，如圖 1-1 的非監督式圖形識別系統所示，我們所要處理的第一個問題為分類的問題，第二個問題為分類個數有效性的問題。根據 2006 年世界盃足球賽 64 場比賽的賽程表 [2] 及圖 1-2 所示，由於進入會內賽共有 32 隊，從比賽開始到

比賽結束分成五個階段共 64 場比賽，我們從這 64 場比賽的原始資料擷取出 32 隊各自的記錄並經資料正規化處理後，分別對全部的球隊進行分類及對分類的統計資料之有效性作驗證。

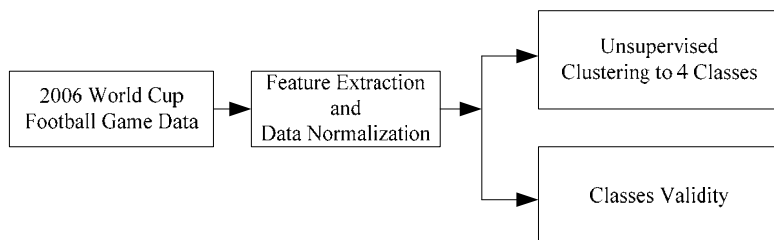
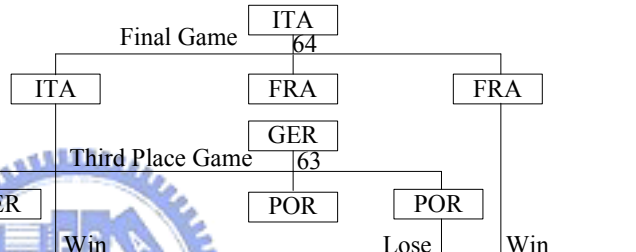


圖 1-1. 非監督式圖形識別於 2006 年世界盃足球賽球隊的分類及分類有效性的系統。

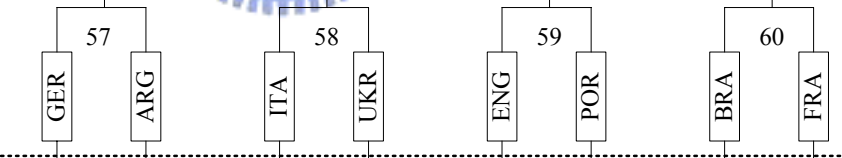
Stage 5 : Finals



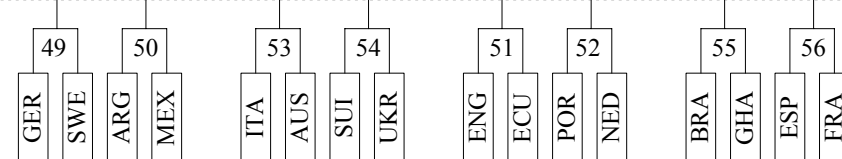
Stage 4 : Semi-Finals



Stage 3 : Quarter-Finals



Stage 2 : Round of 16



Stage 1 : Group Match

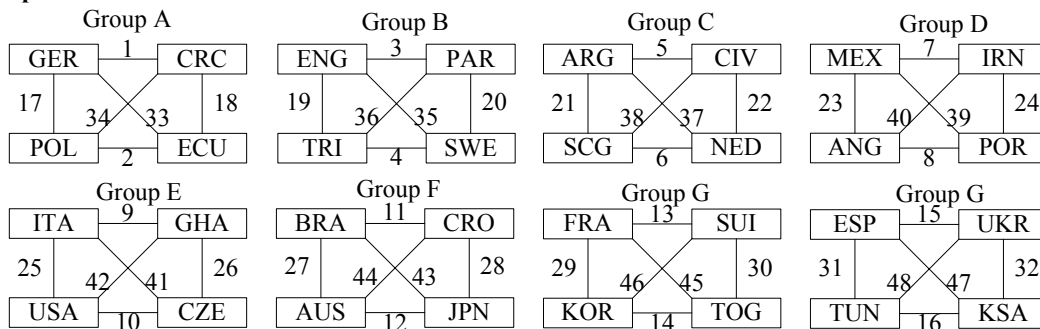


圖 1-2. 2006 年世界盃足球賽 64 場賽程表。

群聚分類方法甚多，在相關學術發表及著作有很多討論 [4] - [6]，根據 Rui Xu 及 Donald Wunsch II 於 2005 年 IEEE 上發表一篇針對各種分類理論法及分類有效性的驗證法 [7]，我們採取了非監督式分類法如圖 1-3 所示，有：K-Means 分類演算法、Fuzzy C-Means 分類演算法、Hierarchical 分類演算法、及 Self-Organizing Feature Map (SOFM) 類神經網路，經由這些分類法來將 2006 年世界盃足球會內賽的 32 個隊伍進行分類，區分哪些隊伍應該是屬於同一類別，其也就是相對的找出強隊族群或弱隊族群，然後比較「分類結果」與「比賽結果」的分類正確率。此外針對 2006 年世界盃足球賽的比賽數據分類數作有效性的分析，我們採取如圖 1-4 所示的 3 種分類有效性的方法，其為 Basic Sequential Cluster Criterion、Modified Basic Sequential Cluster Criterion、及 Pseudo F-Statistic Criterion。

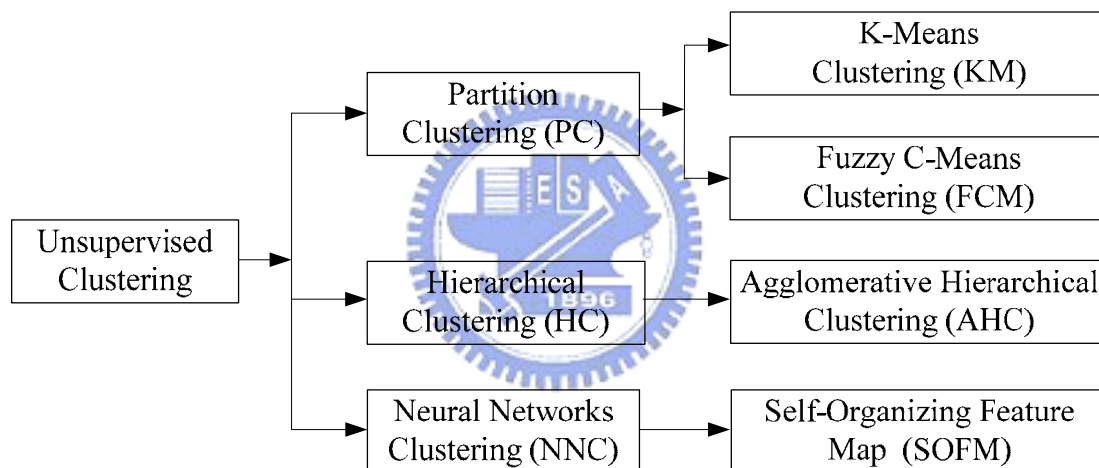


圖 1-3. 所採用的非監督式分類法.

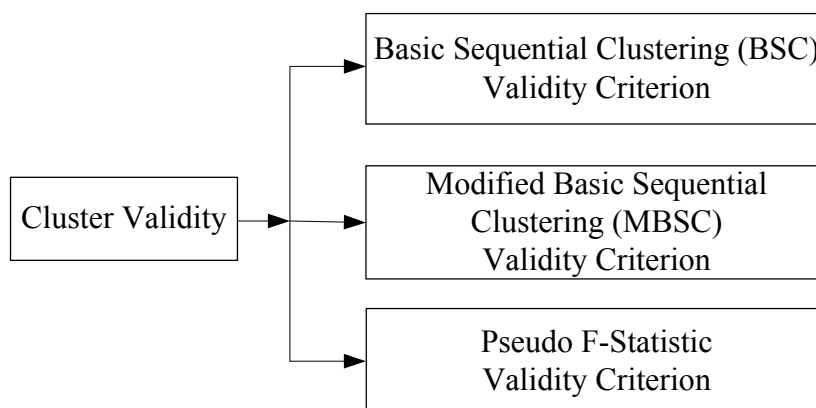


圖 1-4. 所採用的分類有效性驗證法.

## 2. 採用的非監督式分類法

本論文中所採用的非監督式分類法有：K-Means Clustering、Fuzzy C-Means Clustering、Hierarchical Clustering、及 SOFM Clustering，實驗的程式語言為 MATLAB 7.0 版，其中 Fuzzy C-Means Clustering 利用 MATLAB Fuzzy Logic Toolbox 中的 FCM 函數，Hierarchical Clustering 利用 MATLAB Statistics Toolbox 中的 pdist、linkage、及 dendrogram 函數，SOFM Clustering 利用 MATLAB Neural Network Toolbox 中的 SOFM 函數，此外 K-Means Clustering 及 Sequential Clustering 則根據演算法寫程式語言而來。關於各種非監督式演算法之簡介，如下各小節所述。

### 2.1 K-Means 分類法

在分割式分類法 (Partition Clustering) 之中，最基本的方法就是所謂的 K-Means (KM) 分類法 [8]，其主要目的是利用距離的計算，將已知維度為  $n$  的  $N$  個樣本集合  $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ ，分成  $K$  類的集合  $\mathbf{P}_1, \mathbf{P}_2, \dots, \mathbf{P}_K$ ， $\mathbf{P}_i \cap \mathbf{P}_j = \phi$ ，for  $i \neq j$ ，且  $\bigcup_i \mathbf{P}_i = \mathbf{X}$ ，我們採 Euclidian 距離，K-Means 分類演算法的作法如下：

#### K-Means Clustering Algorithm [6]

**Input:**  $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ ，where  $\mathbf{x} \in R^n$ ，and clustering number  $K$ 。

**Output:**  $K$  class centers  $\mathbf{C} = \{\mathbf{c}_1, \dots, \mathbf{c}_K\}$ ，and input patterns in each class。

#### Method:

Step (1). 隨機選取  $K$  個資料點，並視之為  $K$  個類別的中心  $\mathbf{C} = \{\mathbf{c}_1, \dots, \mathbf{c}_K\}$ 。

Step (2). 計算每一樣本  $\mathbf{x}_j$  到每一類的中心  $\mathbf{c}_i$  的距離，然後尋找與其最接近的類的中心，並將  $\mathbf{x}_j$  加入該類  $\mathbf{P}_i$ 。

Step (3). 重新計算每一類新的中心。

Step (4). 如果新的類中心等於舊的類中心，則停止。否則，取新的類中心成為目前的類中心，再回到 Step (2)。

## 2.2 Fuzzy C-Mean 分類法

Fuzzy C-Means (FCM) 也稱做Fuzzy K-Means [9]，它是一種K-Means的改良式演算法，其中加入了模糊邏輯觀念，希望能進一步提升分類的效果，FCM與KM最大的差異在於，KM分類法中，每一個輸入向量僅歸於某一個特定的類別，但是FCM分類法中的每一個輸入向量可隸屬於每一個類別，隸屬的程度稱為membership value  $u_{ij}$ ，其值需滿足式 (2-1) 及式 (2-2)。

$$0 < u_{ij} < 1, \quad i = 1, \dots, C, \quad j = 1, \dots, N \quad (2-1)$$

$$\sum_{i=1}^C u_{ij} = 1, \quad j = 1, \dots, N \quad (2-2)$$

若目標函數  $J$  定義為：

$$J = \sum_{i=1}^C J_i = \sum_{i=1}^C \sum_{j=1}^N u_{ij}^m \| \mathbf{x}_j - \mathbf{m}_i \|^2 \quad (2-3)$$

其中  $m$  為權重係數，可以是介於  $[1, \infty)$  之間的任何數， $\| \mathbf{x}_j - \mathbf{m}_i \|$  為  $\mathbf{x}_j$  到  $\mathbf{m}_i$  的距離函數， $\mathbf{m}_i$  是第  $i$  類的中心。

為了求得最佳化的  $J$ ，可利用Lagrange multiplier method，在滿足式(2-1)及式(2-2)的條件下，針對各傳入參數  $u_{ij}$  與  $\mathbf{m}_i$  分別進行偏微分(對  $\mathbf{m}_i$  則固定  $u_{ij}$ ，對  $u_{ij}$  則固定  $\mathbf{m}_i$ )，並令之為0，整理即可得到：

$$\mathbf{m}_i = \frac{\sum_{j=1}^N u_{ij}^m \mathbf{x}_j}{\sum_{j=1}^N u_{ij}^m}, \quad i = 1, \dots, C \quad (2-4)$$

$$u_{ij} = \frac{\left( \frac{1}{\| \mathbf{x}_j - \mathbf{m}_i \|^2} \right)^{1/(m-1)}}{\sum_{i=1}^C \left( \frac{1}{\| \mathbf{x}_j - \mathbf{m}_i \|^2} \right)^{1/(m-1)}}, \quad i = 1, \dots, C, \quad j = 1, \dots, N \quad (2-5)$$



## Fuzzy C-Means Clustering Algorithm

已知有  $N$  個輸入資料、維度為  $n$  的樣本集合  $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ ，欲將  $\mathbf{x}$  分成  $C$  類，距離計算採 Euclidian 距離，目標函數為  $J$ ，權重矩陣為  $U$ 。

**Input:**  $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ , where  $\mathbf{x} \in R^n$ ;  $C$  classes, and  $\theta$  a threshold.

**Output:** Centers of classes,  $\mathbf{C} = \{\mathbf{c}_1, \dots, \mathbf{c}_c\}$ , where  $\mathbf{c}_i \in R^n$ ; and input data in each class.

**Method:**

Step (1). 設定分類個數為  $C$ ，設定初始權重矩陣  $U$ ，各行列位置的 membership value  $u_{ij}$  須滿足式 (2-1) 及式 (2-2)。

Step (2). 根據式 (2-4) 計算各類別的中心  $\mathbf{m}_i$ ，及根據式(2-5)計算  $u_{ij}$ 。

Step (3). 根據式 (2-3) 計算目標函數  $J$ ，及計算  $\delta = |J(t) - J(t-1)|$ ，當  $\delta \leq \theta$  或疊代次數到達設定的最大次數時，則結束本演算法，否則回到 Step (2)。

## 2.3 凝聚階層式分類法(Agglomerative Hierarchical Clustering (AHC))



Hierarchical Clustering 是根據 Proximity Matrix 的計算把所有資料組織成爲階層的結構來表示類別與類別之間的關係，其中 Proximity Matrix 可爲 Similarity Matrix 或是 Dissimilarity Matrix，這種 Hierarchical Clustering 的結構一般以樹狀圖 (Dendrogram) 來表示，在樹狀圖的樹根節點 (Root Node) 代表的是全部的資料集合，而樹葉節點 (Leaf Node) 代表的是每一個輸入資料，中間連接節點 (Intermediate Node) 代表的是兩個節點(樹根或樹葉節點)之間的連接關係，而樹狀圖的高度通常代表的是類別與類別之間或樣本到類別之間的關係，在樹狀圖中設定一個切割準位即可獲得分類結果，根據不同的切割準位其分類的結果也將不同。本論文中的 Hierarchical Clustering 採用的是 Agglomerative Hierarchical，它的作法是一開始每一個資料本身均代表一個類別，然後藉由將相近的資料合併成爲一群，一直合併直到類別數爲我們所需要的個數爲止，或是把全部的資料集合至最後成爲一個類別後才結束。

若我們定義新組成的群聚  $C_q$  (由  $C_i, C_j$  組成) 與舊群組  $C_s$  之間的距離函數通式可寫成

$$d(C_q, C_s) = a_i d(C_i, C_s) + a_j d(C_j, C_s) + b d(C_i, C_j) + c |d(C_i, C_s) - d(C_j, C_s)| \quad (2-7)$$



則根據不同的  $a_i, a_j, b, c$  的組合設定 [4]，一般又可衍生出下列幾種連結法：

- (1) 單一連結法 (Single Link Algorithm)，也稱做 Nearest Neighbor，如果  $a_i = 1/2, a_j = 1/2, b = 0, c = -1/2$ ，則  $d(C_q, C_s) = \min\{d(C_i, C_s), d(C_j, C_s)\}$ 。
- (2) 完全連結法 (Complete Link Algorithm)，也稱做 Furthest Neighbor，如果  $a_i = 1/2, a_j = 1/2, b = 0, c = -1/2$ ，則  $d(C_q, C_s) = \max\{d(C_i, C_s), d(C_j, C_s)\}$ 。
- (3) 平均連結法 (Average Link Algorithm)，也稱為 Unweighted Pair Group Method Average，是取  $C_i, C_s$  兩個群聚之間所有成員的平均距離與  $C_j, C_s$  兩個群聚之間所有成員的平均距離較小者成為新的群聚  $C_q$ ，新組成的群聚  $C_q$  與舊群聚  $C_s$  的距離公式就如式 2-8 所示，其中  $n_i$  為  $i$  群聚中的成員數，而  $n_j$  為  $j$  群聚中的成員數。如果

$$a_i = \frac{n_i}{n_i + n_j}, a_j = \frac{n_j}{n_i + n_j}, b = 0, c = 0, \text{ 則}$$

$$d(C_q, C_s) = \frac{n_i}{n_i + n_j} d(C_i, C_s) + \frac{n_j}{n_i + n_j} d(C_j, C_s) \quad (2-8)$$



## 2.4 Self-Organizing Feature Map Neural Networks (SOFM) 分類法

Self-Organizing Feature Map (SOFM) [10]-[14] 網路架構由芬蘭的 Kohonen 於 1980 年提出，如圖 2-1 所示，它是一種兩層、順向連結的前饋式競爭學習類神經網路，主要元件包括輸入層、輸出層、及連結單元。連結單元為輸出層神經元與輸入層神經元之間相互連結的加權值所構成的向量，而輸出神經元會安排成前後左右有意義的空間關係稱為拓撲結構 (topological structure)，不同形式空間位置安排的拓撲結構及鄰近區域的觀念組成的 SOFM 也稱為拓撲圖 (topology)，如圖 2-2 所示，即為常用到的 3 種不同 Topology，其中圖 2-2(a) 所示，其輸出神經元在空間上排列成一直線，稱為 1D Topology，圖 2-2(b) 所示，其輸出神經元在空間排成 2D 的格子狀，稱為 2D Grid Topology，圖 2-2(c) 所示，其輸出神經元在空間排成 2D 的六角形稱為 2D Hexagonal Topology，關於 SOFM 的演算法說明可參閱 [6]。

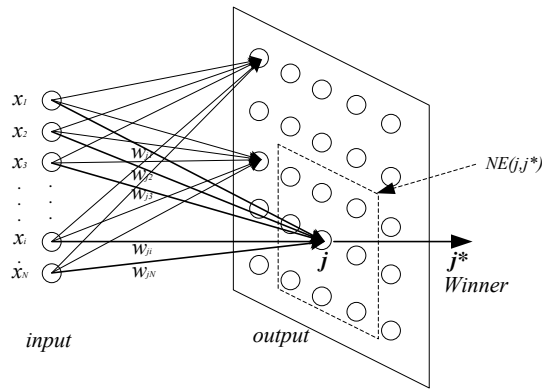


圖 2-1. SOFM 網路架構.

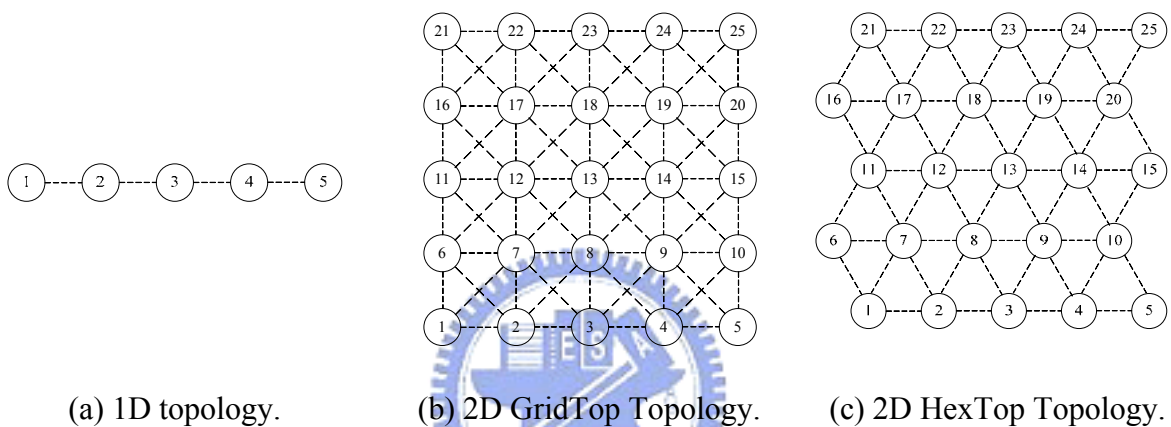


圖 2-2. 3 種常用的輸出神經元的 Topology.

## 2.5 順序性分類演算法 (Sequential Clustering Algorithm)

Sequential Clustering Algorithm (SC) 由「Hall A.V.」在1967年提出 [15]，它是一種非常簡單且快速的非監督式分類法，這個演算法中的每一個樣本向量只會出現1次或是很少的次數(一般來說不超過5次)，分類最終數目一開始時並不知道，使用者必需設定一個門檻值 $\theta$ ，這個門檻值 $\theta$ 用來做為合併或獨立成一個類別的判斷基準，另外必需設定一個最大分類 $U$ ，而一開始的類別只有一群，然後隨著一個接一個的樣本向量送進來分類後，樣本會被合併到之前的群集中，或是獨立自成一個群集，直到樣本分配完畢後結束，更詳細的說明可參閱 [4]，由於下一節的分類個數有效性驗證會利用到Sequential Clustering Algorithm，因此以下小節將針對Basic Sequential Clustering (BSC) 及Modified Basic Sequential Clustering (MBSC) 作說明。

## (1) Basic Sequential Clustering (BSC)

若  $d(\mathbf{x}, P)$  代表樣本  $\mathbf{x}$  與類別  $P$  的距離， $P$  的類中心點為  $\mathbf{M}$ ，所以  $d(\mathbf{x}, P) = d(\mathbf{x}, \mathbf{M})$ ，當有一個新樣本  $\mathbf{x}$  加到離其最近的  $P$  類後，該類的中心  $\mathbf{M}$  需用式 (2-5) 來更新，其中  $N_{\mathbf{M}_{closest}}^{new}$  為加入成員後的樣本個數。

$$\mathbf{M}_{closest}^{new} = \frac{(N_{\mathbf{M}_{closest}}^{new} - 1)\mathbf{M}_{closest}^{old} + \mathbf{x}}{N_{\mathbf{M}_{closest}}^{new}} \quad (2-5)$$

初始條件需給一個 distance threshold,  $\theta$ , 及最大允許生成類別數  $U$ 。當一個樣本進來時，先找到離其最近的類別之距離，當距離大於  $\theta$  時就將樣本獨立成一個類別，當距離小於  $\theta$  時，就把它合併到距離其最近的那個類別，然後更新該類的中心  $\mathbf{M}$ ，直到生成的類別數達到最大生成數  $U$  時就結束。

### Basic Sequential Clustering Algorithm (BSC) [4]

**Input :** Given  $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ , distance threshold  $= \theta$ , and maximum allowed cluster number  $U$ .

**Output :** Output data in each of  $K$  classes.

**Method :**

Step (1).  $K=1$ ,  $P_K = \{\mathbf{x}_1\}$

Step (2). For  $\mathbf{x}_i$ ,  $i = 2, \dots, N$

(2.1) Find  $P_{closest}$ , where  $d(\mathbf{x}_i, P_{closest}) = \min_{1 \leq j \leq K} d(\mathbf{x}_i, \mathbf{M}_j)$

(2.2) If  $(d(\mathbf{x}_i, P_{closest}) > \theta)$  and  $(K < U)$ , then

$K=K+1$ ;

$P_K = \{\mathbf{x}_i\}$

else

$P_{closest} = \{P_{closest}\} \cup \{\mathbf{x}_i\}$ ;

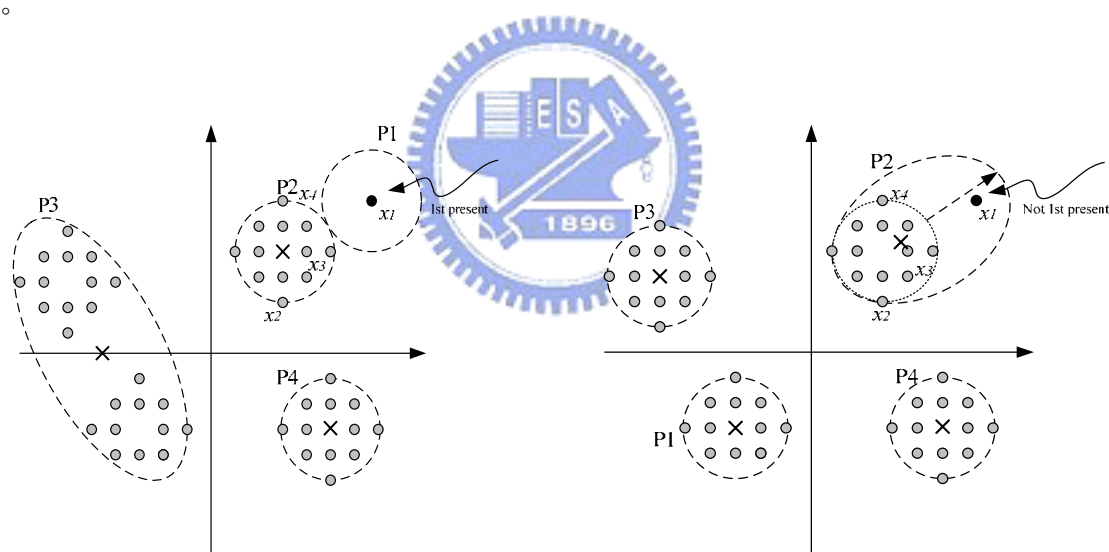
Update mean vector  $\mathbf{M}_{closest}$

End

End

## (2) Modified Basic Sequential Clustering (MBSC)

由於 BSC 演算法中，對輸入樣本的分配採先進來就先處理，所以對於原本屬於同一類別內同性質的樣本來說，當  $x_1$  是第一個出現時，會先被獨立成一群稱  $P_1$ ，接著  $x_2$  進來時若由於距離高於門檻值  $\theta$  而被獨立成一群稱  $P_2$ ，接著  $x_3$  及其之後的樣本陸續進來而被合併到  $P_2$  群，由於群中心需即時更新，若是  $P_2$  類別的版圖明顯地向  $P_1$  擴張到原本應把  $x_1$  要包含進去，但礙於  $x_1$  已經先出現過且被獨立成一群  $P_1$ ，所以無法及時更正這樣的錯誤，如圖 2-3(a) 所示即為此狀況下經 BSC 分類的結果，另外一個例子如圖 2-3(b) 所示，其樣本  $x_1$  非第一次出現，然後經 BSC 分類的結果。所以本節的 MSBC 就是採取兩段式作法，第一階段是讓所有的樣本先出現過一次，大致上先決定欲分類的個數之後，第二階段再把所有樣本中尚未被分類的樣本重新分配至第一階段所產生出來的類別中。



(a) 樣本  $x_1$  為第一個出現之分類的結果。 (b) 樣本  $x_1$  非第一個出現之分類的結果。

圖 2-3. 同一個樣本，在不同時間出現，經 BSC 分類的結果。

## Modified Basic Sequential Clustering Algorithm (MBSC) [4]

**Input :** Given  $X = \{x_1, x_2, \dots, x_N\}$ , distance threshold  $= \theta$ , and maximum allowed cluster number  $U$ .

**Output :** Output data in each of  $K$  classes.

**Method :**

**Phase I. (決定分類數)**

Step (1).  $K=1$ ,  $P_K = \{x_1\}$

Step (2). For  $x_i$ ,  $i=2, \dots, N$

(2.1) Find  $P_{closest}$ , where  $d(x_i, P_{closest}) = \min_{1 \leq j \leq K} d(x_i, M_j)$

(2.2) If  $(d(x_i, P_{closest}) > \theta)$  and  $(K < U)$ , then

$K=K+1$ ;

$P_K = \{x_i\}$

End

End

(continue)

**Phase II. (樣本分類)**

Step (3). For  $x_i$ ,  $i=1, \dots, N$

(3.1) If  $x_i$  has not been assigned to a cluster, then

Find  $P_{closest}$ , where  $d(x_i, P_{closest}) = \min_{1 \leq j \leq K} d(x_i, M_j)$

$P_{closest} = \{P_{closest}\} \cup \{x_i\}$ ;

Update mean vector  $M_{closest}$

End

End



### 3. 採用的分類有效性驗證法

一般在分類應用上，我們可先假設資料本身具有分類的特性，然後在分類結束時除了解釋分類的結果之外，我們可以針對原始資料再進行驗證的動作，從驗證的結果我們可以探索分類的數量是否合理或理想，我們稱這樣的驗證過程為分類有效性的評估 [4]。

分類結果的好與壞，有兩種判定的方式：

- (1) 對同一類別內成員的關係可用相似度來解釋，被分到同一類的成員與成員彼此之間愈靠近，其相似度測量值就愈高，它代表同一類別內，成員彼此之間的聚集程度愈緊密，愈緊密的程度代表的就是比較好的分類結果，例如：

第  $P_i$  類的 scatter matrix 定義為：

$$S_i = \sum_{x \in P_i} (\mathbf{x} - \mathbf{M}_i)(\mathbf{x} - \mathbf{M}_i)^T \quad (3-1)$$

其中  $\mathbf{M}_i$  為第  $P_i$  類的類中心，定義  $\mathbf{M}_i = \frac{1}{n_i} \sum_{x \in P_i} \mathbf{x}$ 。

對全部類別的 within-cluster scatter matrix 定義為：

$$S_W = \sum_{i=1}^K S_i \quad (3-2)$$

我們定義  $trS_W$  為：

$$trS_W = \sum_{i=1}^K trS_i = \sum_{i=1}^K \sum_{x \in P_i} \|\mathbf{x} - \mathbf{M}_i\|^2 \quad (3-3)$$

愈小的  $trS_W$  值，代表同一類別的成員聚集程度愈緊密，表示較好的分類效果。

- (2) 對不同類別與類別之間的關係則可用分離度來解釋，類別與類別彼此之間愈分離就代表兩個類別的分離度愈大，其差異性也愈大，所以應該被分的更遠，因此分離度愈大代表的是比較好的分類結果，例如：

對全部類別的 between-cluster scatter matrix 定義為：

$$S_B = \sum_{i=1}^K n_i (\mathbf{M}_i - \mathbf{M})(\mathbf{M}_i - \mathbf{M})^T \quad (3-4)$$

其中  $\mathbf{M}$  為 global mean，定義為  $\mathbf{M} = \frac{1}{N} \sum_{x \in P} \mathbf{x} = \frac{1}{N} \sum_{i=1}^K n_i \mathbf{M}_i$ 。

我們定義  $trS_B$  為：

$$trS_B = \sum_{i=1}^K n_i \|\mathbf{M}_i - \mathbf{M}\|^2 \quad (3-5)$$

愈大的  $trS_B$  值，代表不同類別與類別之間的距離愈遠，表示較好的分類效果。

本論文應用於 2006 世界盃足球賽統計數據之分類個數有效性驗證法有三種，分別為以 Sequential Clustering 為基礎的 Basic Sequential Clustering (BSC) 及 Modified Basic Sequential Clustering (MBSC) 兩種，以及以 K-Means 為基礎的 Pseudo F-Statistics (PFS) 一種，此三種有效性驗證方法分別詳述於下面各小節，而實驗程式則根據公式及演算法利用 MATLAB 7.0 實作而來。

### 3.1 以 BSC 及 MBSC 為基礎的有效性驗證法

#### (1) BSC Validity Criterion



若輸入資料為  $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ ，我們計算找出樣本間最小距離  $a = \min_{i,j=1,\dots,N} d(x_i, x_j)$ ，及樣本間最大距離  $b = \max_{i,j=1,\dots,N} d(x_i, x_j)$ ，然後門檻值  $\theta$  由最小距離  $a$  開始測試，每一個門檻值  $\theta$  設定值共測試  $s$  次 BSC 演算法，然後找出該門檻值在  $s$  次測試當中最常出現的  $K$  值，並記錄下來，然後將門檻值  $\theta$  加上增量  $c$  成為新的門檻值  $\theta$ ，重覆先前步驟直到門檻值  $\theta$  到達最大距離  $b$  就結束。由於  $\theta$  的值將直接影響到樣本與類別中心計算後歸類的決定，因此  $\theta$  的精確度選擇相當重要， $s$  代表每一個門檻測試值需要被測試的次數，愈多次的疊代測試所得到的統計結果愈精確，此外每一次疊代測試時樣本輸入的順序皆需改變。

經由上述的方法，我們可以得到一個門檻值  $\theta$  與  $K$  相對應的二維空間分佈曲線圖，由分佈圖中的曲線我們找出當隨著門檻值  $\theta$  緩慢變化時，其相對應的  $K$  值若有出現顯著的變化轉折 (Knee)，則這樣的轉折變化即代表輸入資料擁有類別的特性，而在此曲線中，若隨著門檻值  $\theta$  緩慢增加時，其所相應的  $K$  值若維持不變 ( $K$  值維持平坦的區域)，則該  $K$  值即可成為分類的個數  $K$  [4]。



利用 Basic Sequential Clustering Algorithm 找尋分類個數  $K$  值的演算法如下：

### BSC Validity Criterion Algorithm

**Input:**  $\mathbf{X} = \{x_1, \dots, x_N\}$ ,  $a = \min_{i,j=1,\dots,N} d(x_i, x_j)$ ,  $b = \max_{i,j=1,\dots,N} d(x_i, x_j)$ ,  $\theta$  (distance threshold),  $c$  (step size of  $\theta$ ),  $s$  (the number of times for each  $\theta$ ).

**Output:** The most frequent  $K$  for each  $\theta$ .

**Method:**

Step (1).  $j=0$

Step (2). For ( $\theta = a$ ;  $\theta < b$ ;  $\theta = \theta + c$ ) {

$i=0$ ;

While ( $i < s$ ) {

$U[i] = BSC(\mathbf{x}, \theta)$

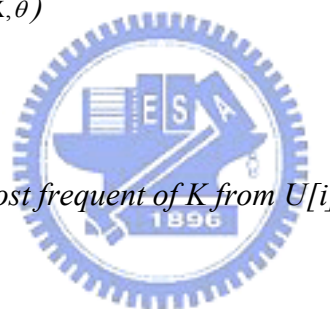
$i = i + 1$ ;

} //end of While

$K[j] = \text{find the most frequent of } K \text{ from } U[i]$

$j = j + 1$ ;

} //end of For



Step (3). Plot  $K$  versus  $\theta$ .

### (2) MBSC Validity Criterion

利用改良形順序性分類法(Modified Basic Sequential Clustering)的方法來找尋分類個數  $K$  值的演算法與上述方法相同，差別在本節所使用的演算法為 MBSC。

### 3.2 以 K-Means 為基礎的有效性驗證法：Pseudo F-Statistics Validity Criterion

Vogel and Wong 於 1979 年 [16] 提出 Pseudo F-Statistics (PFS)，它是以 K-Means 為基礎的有效性驗證法，由最大 F-Statistics 的值來決定分類數。其中 PFS 公式定義如下：

$$PFS = \frac{\frac{tr\mathbf{S}_B}{K-1}}{\frac{tr\mathbf{S}_W}{N-K}} = \frac{tr\mathbf{S}_B * (N-K)}{tr\mathbf{S}_W * (K-1)} \quad (3-6)$$

$N$  為樣本的總數， $K$  為分類的數目， $tr\mathbf{S}_W = \sum_{i=1}^K tr\mathbf{S}_i = \sum_{i=1}^K \sum_{\mathbf{x} \in P_i} \|\mathbf{x} - \mathbf{M}_i\|^2$ ， $tr\mathbf{S}_B = \sum_{i=1}^K n_i \|\mathbf{M}_i - \mathbf{M}\|^2$ 。

由式 (3-6) 的計算中，我們可以找到一個 PFS 的 Peak 值，這個 Peak 值對應的  $K$  值的分類結果代表的就是同一類別具有相對比較緊密的群聚效果(較小的  $tr\mathbf{S}_W$ )，及不同類別之間具有比較分離的分類效果(較大的  $tr\mathbf{S}_B$ )。



## 4. 非監督式分類法用於 2006 年世界盃足球賽球隊實力的分類

### 4.1 原始資料的擷取及轉換

根據 2006 年世界盃足球賽的官方網頁 (<http://fifaworldcup.yahoo.com>) 我們可得到每一場比賽的比賽報告，如附錄二所示為德國(GER)對哥斯大黎加(CRC)之比賽報告，共計有 64 場，每場比賽報告的統計項目共計有 17 項，如表 4-1 所列，從這 17 個統計項目中，我們擷取其中可以代表球隊實力或影響球隊實力的 8 項統計數據做為輸入資料的特徵值，分別標示為  $x_1 \sim x_8$ ， $x_1$  為進球數 (Goals For (GF))， $x_2$  為射門次數 (Shots (S))， $x_3$  為射門命中球門次數 (Shots On Goal (SOG))， $x_4$  為角球數 (Corner Kicks (CK))， $x_5$  為直接自由球命中球門數 (Direct Free Kicks to Goal (DFKG))， $x_6$  為間接自由球命中球門數 (Indirect Free Kicks to Goal (IDFG))， $x_7$  為持球時間 (Ball Possession (BP))， $x_8$  為被犯規次數 (Fouls Suffered (FS))，這 8 項特徵值在輸入分類演算法之前須先經過正規化處理，其中  $x_1 \sim x_8$  的值愈大代表球隊實力愈強，其輸入值的正規化計算為：

*If*  $x_{iA} = x_{iB}$ , *then*

$$y_{iA} = y_{iB} = 0.5$$

*Else*

$$y_{iA} = \frac{x_{iA}}{x_{iA} + x_{iB}}, \quad y_{iB} = \frac{x_{iB}}{x_{iA} + x_{iB}}, \quad i = 1, \dots, 8$$

$A$  代表  $A$  隊， $B$  代表  $B$  隊， $i$  為 8 個特徵值的索引，上述正規化計算後， $x_{iA}$  與  $x_{iB}$  的值將介於 0~1 之間。我們設 *If*  $x_{iA} = x_{iB}$ , *then*  $y_{iA} = y_{iB} = 0.5$ ，是因為如果  $A$  隊及  $B$  隊的進球數均為 0，則輸入值變為  $y_{iA} = 0.5$ ,  $y_{iB} = 0.5$ 。

德國(GER)對哥斯大黎加(CRC) 8 項原始資料的擷取及正規化轉換後的資料列於表 4-2。

表 4-1. 德國隊(GER)對哥斯大黎加隊(CRC) 17 項原始資料.

(技術指標)統計項目	簡寫	選為特徵值	A 球隊(GER)	B 球隊(CRC)
進球數 (Goals For)	GF	$x_1$	4	2
失球數 (Goal Against)	GA		2	4
射門次數 (Shots)	S	$x_2$	21	4
射門命中球門數 (Shots On Goal)	SOG	$x_3$	10	2
十二碼罰球數 (Penalty Kicks)	PK		0	0
犯規次數 (Fouls Committed)	FC		11	15
黃牌數 (Yellow Cards)	Y		0	1
紅牌數 (Red Cards)	R		0	0
角球數 (Corner Kicks)	CK	$x_4$	7	3
直接自由球命中球門數 (Direct Free Kicks to Goal)	DFKG	$x_5$	1	0
間接自由球命中球門數 (Indirect Free Kicks to Goal)	IDFKG	$x_6$	0	0
越位次數 (Offside)	O		3	3
烏龍球 (Own Goals)	OG		0	0
警告 (Cautions)	C		0	1
驅逐出場 (Expulsions)	E		0	0
持球時間 (Ball Possession)	BP	$x_7$	63%	37%
被犯規次數 (Fouls Suffered)	FS	$x_8$	12	11

Note: 1. 十二碼球次數只含正規時間及加時賽的部份，PK 大戰不列入統計。

2. GER 的「被犯規次數」與 CRC 的「犯規次數」不同，是因為犯規次數的記錄含非對人的犯規，例如：手球，邊線球。

表 4-2. 德國隊(GER)對哥斯大黎加隊(CRC) 8 項原始資料的擷取及正規化轉換.

(技術指標)統計項目	簡寫	特徵值	A 球隊(GER) 轉換前 $x_{iA}$	B 球隊(CRC) 轉換前 $x_{iB}$	A 球隊(GER) 轉換後 $y_{iA}$	B 球隊(CRC) 轉換後 $y_{iB}$
Goals For	GF	$x_1$	4	2	0.6666	0.3333
Shots	S	$x_2$	21	4	0.84	0.16
Shots On Goal	SOG	$x_3$	10	2	0.8333	0.1666
Corner Kicks	CK	$x_4$	7	3	0.7	0.3
Direct Free Kicks to Goal	DFKG	$x_5$	1	0	1	0
Indirect Free Kicks to Goal	IDFKG	$x_6$	0	0	0.5	0.5
Ball Possession	BP	$x_7$	63%	37%	0.63	0.37
Fouls Suffered	FS	$x_8$	12	11	0.5217	0.4782

從 2006 年世界盃足球賽的 64 場紀錄資料中擷取出來的原始資料，如表 4-3 所示，會先經過上述之正規化轉換後成為表 4-4，接著我們分別把 32 個球隊各自經轉換後的資料全部加起來後得到 32 隊的各自資料的總和，如表 4-5 所示，由於各隊的總比賽場次 (Match Played (MP)) 並不相同，因此各隊會先除以自己的比賽場次後得到 32 隊各隊的單場平均輸入資料，如表 4-6 所示。

表 4-3. 32 隊全部比賽的原始資料.

Group	Team	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$	$x_8$	Group	Team	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$	$x_8$	
		GF	S	SG	CK	DFKG	IDFKG	BP	FS			GF	S	SG	CK	DFKG	IDFKG	BP	FS	
A	<b>GER</b>									E	<b>ITA</b>									
	Match 1	4	21	10	7	1	0	63	12		Match 9	2	18	13	12	0	2	47	21	
	Match 17	1	16	8	10	0	0	58	17		Match 25	1	10	3	7	1	0	54	22	
	Match 33	3	15	9	2	0	0	43	21		Match 41	2	14	6	5	1	0	52	16	
	Match 49	2	26	11	4	0	0	63	20		Match 53	1	11	6	2	1	0	41	24	
	Match 57	1	10	5	4	2	1	42	30		Match 58	3	10	7	1	1	0	41	26	
	Match 61	0	13	2	4	1	0	43	19		Match 61	2	15	10	12	1	0	57	19	
	Match 63	2	12	5	2	2	1	43	14		Match 64	1	5	3	5	1	0	55	24	
	<b>CRC</b>										<b>GHA</b>									
	Match 1	2	4	2	3	0	0	37	11		Match 9	0	14	4	4	0	0	53	8	
	Match 18	0	12	4	4	0	0	49	18		Match 26	2	20	8	7	0	0	50	16	
	Match 34	1	12	5	2	3	0	49	20		Match 42	2	9	4	2	0	1	48	15	
	<b>POL</b>										Match 55	0	18	7	4	1	0	52	16	
	Match 2	0	7	3	11	1	1	56	15		<b>USA</b>									
	Match 17	0	5	3	4	0	0	42	21		Match 10	0	6	1	2	0	0	55	17	
Match 34	2	10	7	8	2	0	51	12	Match 25	0	8	0	3	1	0	46	13			
<b>BCU</b>									Match 42	1	7	3	7	0	0	52	30			
Match 2	2	10	6	2	1	0	44	9	<b>CZE</b>											
Match 18	3	14	7	3	2	0	51	22	Match 10	3	10	5	5	0	0	45	14			
Match 33	0	7	2	5	2	0	57	18	Match 26	0	14	4	6	1	1	50	22			
Match 51	0	9	3	7	0	0	49	13	Match 41	0	11	8	4	0	0	48	15			
B	<b>ENG</b>									F	<b>BRA</b>									
	Match 3	0	13	5	6	0	0	53	11		Match 11	1	13	6	5	0	0	50	19	
	Match 19	2	23	8	7	0	0	62	19		Match 27	2	16	6	7	2	0	54	24	
	Match 35	2	14	8	6	1	0	55	17		Match 43	4	21	14	11	1	0	60	9	
	Match 51	1	8	4	5	2	0	49	22		Match 55	3	11	10	3	1	0	48	22	
	Match 59	0	9	4	6	1	0	43	10		Match 60	0	7	1	5	2	0	55	17	
	<b>PAR</b>										<b>CRO</b>									
	Match 3	0	7	2	1	0	0	47	12		Match 11	0	9	3	7	1	0	50	19	
	Match 20	0	16	3	3	1	0	43	18		Match 28	0	16	6	11	0	0	44	19	
	Match 36	1	16	9	7	1	0	53	19		Match 44	2	8	3	4	1	0	44	23	
	<b>TRI</b>										<b>AUS</b>									
	Match 4	0	6	2	1	0	0	40	7		Match 12	3	20	12	5	3	0	52	11	
	Match 19	0	7	3	3	0	0	38	14		Match 27	0	14	4	4	0	0	46	8	
	Match 36	0	9	2	1	0	0	47	16		Match 44	2	12	7	9	0	0	56	20	
	<b>SWE</b>										Match 53	0	8	4	2	1	0	59	17	
Match 4	0	18	6	8	1	0	60	10	<b>JPN</b>											
Match 20	1	17	10	6	1	0	57	15	Match 12	1	6	2	3	0	0	48	22			
Match 35	2	9	6	12	1	0	45	11	Match 28	0	12	5	5	3	0	56	18			
Match 49	0	5	2	4	0	0	37	15	Match 43	1	9	3	3	1	0	40	5			
C	<b>ARG</b>									G	<b>PRA</b>									
	Match 5	2	9	4	3	2	0	49	17		Match 13	0	9	3	4	0	0	51	16	
	Match 21	6	11	9	3	0	0	58	20		Match 29	1	15	4	6	0	0	52	10	
	Match 37	0	10	3	10	0	0	47	21		Match 45	2	17	9	9	0	0	56	21	
	Match 50	2	11	5	6	0	0	51	25		Match 56	3	9	5	5	1	0	39	27	
	Match 57	1	12	5	6	1	0	58	23		Match 60	1	9	5	7	1	0	45	21	
	<b>CIV</b>										Match 62	1	5	4	3	0	0	41	16	
	Match 5	1	13	4	6	0	1	51	15		Match 64	1	13	5	7	0	0	45	15	
	Match 22	1	16	9	8	0	1	49	23		<b>SUI</b>									
	Match 38	3	20	10	9	0	0	68	17		Match 13	0	7	4	1	1	1	49	18	
	<b>SCG</b>										Match 30	2	15	9	8	0	0	50	17	
	Match 6	0	11	4	6	2	1	39	22		Match 46	2	12	6	8	1	0	46	19	
	Match 21	0	4	1	4	0	0	42	13		Match 54	0	12	6	5	3	0	55	19	
	Match 38	2	6	3	1	2	0	32	12		<b>KOR</b>									
	<b>NBD</b>										Match 14	2	16	6	3	2	0	64	17	
Match 6	1	12	6	4	1	0	61	14	Match 29	1	5	2	2	2	0	48	20			
Match 22	2	9	8	3	1	0	51	15	Match 46	0	15	8	6	0	0	54	7			
Match 37	0	9	3	7	2	1	53	16	<b>TOG</b>											
Match 52	0	20	9	5	2	0	62	8	Match 14	1	9	3	4	1	1	36	15			
D	<b>MEX</b>									H	<b>ESP</b>									
	Match 7	3	7	4	6	1	0	53	18		Match 15	4	19	10	7	2	0	54	14	
	Match 23	0	13	8	6	2	0	54	20		Match 31	3	24	10	12	2	0	66	22	
	Match 39	1	14	6	5	1	0	50	27		Match 47	1	19	13	10	0	0	59	20	
	Match 50	1	12	3	5	0	0	49	19		Match 56	1	7	2	8	1	0	61	21	
	<b>IRN</b>										<b>UKR</b>									
	Match 7	1	7	5	5	0	0	47	25		Match 15	0	5	2	1	0	0	46	11	
	Match 24	0	5	1	1	0	0	37	18		Match 32	4	19	9	6	0	0	50	22	
	Match 40	1	18	13	3	1	0	55	23		Match 48	1	9	6	3	1	0	47	23	
	<b>ANG</b>										Match 54	0	10	2	6	1	0	45	22	
	Match 8	0	11	3	2	0	0	42	19		Match 58	0	13	7	3	1	0	59	15	
	Match 23	0	8	1	5	2	0	46	16		<b>TUN</b>									
	Match 40	1	15	7	6	1	0	45	17		Match 16	2	6	2	3	0	0	49	12	
	<b>POR</b>										Match 31	1	4	3	1	1	0	34	8	
	Match 8	1	16	8	5	1	0	58	28		Match 48	0	9	3	3	2	0	53	17	
Match 24	2	18	10	13	2	0	63	18	<b>KSA</b>											
Match 39	2	11	5	4	0	0	50	14	Match 16	2	13	5	4	3	1	51	16			
Match 52	1	10	6	3	1	0	38	12	Match 32	0	6	0	2	0	0	50	23			
Match 59	0	20	9	4	0	0	57	18	Match 47	0	7	4	4	2	0	41	21			
Match 62	0	12	5	8	1	0	59	11												
Match 63	1	13	8	7	3	0	57	15												

GF:Goals For S:Shots SG:Shots On Goal CK:Corner Kicks DFKG:Direct Free Kicks to Goal IDFKG:Indirect Free Kicks to Goal BP: Ball Possession FS:Fouls Suffered

表 4-4. 32 隊原始資料經轉換後的資料.

Group	Team	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>	X <sub>6</sub>	X <sub>7</sub>	X <sub>8</sub>	Group	Team	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>	X <sub>6</sub>	X <sub>7</sub>	X <sub>8</sub>				
		GF	S	SG	CK	DFKG	IDFKG	BP	FS			GF	S	SG	CK	DFKG	IDFKG	BP	FS				
A	GER	Match 1	0.667	0.840	0.833	0.700	1.000	0.500	0.630	0.522	E	ITA	Match 9	1.000	0.563	0.765	0.750	0.500	1.000	0.470	0.724		
	Match 17	1.000	0.762	0.727	0.714	0.500	0.500	0.580	0.447	Match 25		1.000	0.556	1.000	0.700	0.500	0.500	0.500	0.540	0.629			
	Match 33	1.000	0.682	0.818	0.286	0.000	0.500	0.430	0.538	Match 41		1.000	0.560	0.429	0.556	1.000	0.500	0.500	0.520	0.516			
	Match 49	1.000	0.839	0.846	0.500	0.500	0.500	0.630	0.667	Match 53		1.000	0.579	0.600	0.500	0.500	0.500	0.500	0.410	0.585			
	Match 57	0.500	0.455	0.500	0.400	0.667	1.000	0.420	0.566	Match 58		1.000	0.435	0.500	0.250	0.500	0.500	0.500	0.410	0.634			
	Match 61	0.000	0.464	0.167	0.250	0.500	0.500	0.430	0.500	Match 61		1.000	0.536	0.833	0.750	0.500	0.500	0.500	0.570	0.500			
	Match 63	0.667	0.480	0.385	0.222	0.400	1.000	0.430	0.483	Match 64		0.500	0.278	0.375	0.417	1.000	0.500	0.550	0.615				
	CRC	Match 1	0.333	0.160	0.167	0.300	0.000	0.500	0.370	0.478		GHA	Match 9	0.000	0.438	0.235	0.250	0.500	0.000	0.530	0.276		
	Match 18	0.000	0.462	0.364	0.571	0.000	0.500	0.490	0.450	Match 26		1.000	0.588	0.667	0.538	0.000	0.000	0.000	0.500	0.421			
	Match 34	0.333	0.545	0.417	0.200	0.600	0.500	0.490	0.625	Match 42		0.667	0.563	0.571	0.222	0.500	1.000	0.480	0.333				
	POL	Match 2	0.000	0.412	0.333	0.846	0.500	1.000	0.560	0.625		Match 55	0.000	0.621	0.412	0.571	0.500	0.500	0.520	0.421			
	Match 17	0.000	0.238	0.273	0.286	0.500	0.500	0.420	0.553	USA		Match 10	0.000	0.375	0.167	0.286	0.500	0.500	0.550	0.548			
	Match 34	0.667	0.455	0.583	0.800	0.400	0.500	0.510	0.375	Match 25		0.000	0.444	0.000	0.300	0.500	0.500	0.460	0.371				
	ECU	Match 2	1.000	0.588	0.667	0.154	0.500	0.000	0.440	0.375		Match 42	0.333	0.438	0.429	0.778	0.500	0.000	0.520	0.667			
	Match 18	1.000	0.538	0.636	0.429	1.000	0.500	0.510	0.550	CZE		Match 10	1.000	0.625	0.833	0.714	0.500	0.500	0.450	0.452			
	Match 33	0.000	0.318	0.182	0.714	1.000	0.500	0.570	0.462	Match 26		0.000	0.412	0.333	0.462	1.000	1.000	0.500	0.579				
	Match 51	0.000	0.529	0.429	0.583	0.000	0.500	0.500	0.371	Match 41		0.000	0.440	0.571	0.444	0.000	0.500	0.480	0.484				
	B	ENG	Match 3	0.500	0.650	0.714	0.857	0.500	0.500	0.530		0.478	F	BRA	Match 11	1.000	0.591	0.667	0.417	0.000	0.500	0.500	0.500
Match 19		1.000	0.767	0.727	0.700	0.500	0.500	0.620	0.576	Match 27	1.000	0.533		0.600	0.636	1.000	0.500	0.540	0.750				
Match 35		0.500	0.609	0.571	0.333	0.500	0.500	0.550	0.607	Match 43	0.000	0.700		0.824	0.786	0.500	0.500	0.600	0.643				
Match 51		1.000	0.471	0.571	0.417	1.000	0.500	0.500	0.629	Match 55	1.000	0.379		0.588	0.429	0.500	0.500	0.480	0.579				
Match 59		0.500	0.310	0.308	0.600	1.000	0.500	0.430	0.357	Match 60	0.000	0.438		0.167	0.417	0.667	0.500	0.550	0.447				
PAR		Match 3	0.500	0.350	0.286	0.143	0.500	0.500	0.470	0.522	CRO	Match 11		0.000	0.409	0.333	0.583	1.000	0.500	0.500	0.500		
Match 20		0.000	0.485	0.231	0.333	0.500	0.500	0.430	0.545	Match 28	0.500	0.571		0.545	0.688	0.000	0.500	0.440	0.514				
Match 36		1.000	0.640	0.818	0.875	1.000	0.500	0.530	0.543	Match 44	0.500	0.400		0.300	0.308	1.000	0.500	0.440	0.535				
TRI		Match 4	0.500	0.250	0.250	0.111	0.000	0.500	0.400	0.412	AUS	Match 12		0.750	0.769	0.857	0.625	1.000	0.500	0.520	0.333		
Match 19		0.000	0.233	0.273	0.300	0.500	0.500	0.380	0.424	Match 27	0.000	0.467		0.400	0.364	0.000	0.500	0.460	0.250				
Match 36		0.000	0.360	0.182	0.125	0.000	0.500	0.470	0.457	Match 44	0.500	0.600		0.700	0.692	0.000	0.500	0.560	0.465				
SWE		Match 4	0.500	0.750	0.750	0.889	1.000	0.500	0.600	0.588	Match 53	0.000		0.421	0.400	0.500	0.500	0.500	0.590	0.415			
Match 20		1.000	0.515	0.769	0.667	0.500	0.500	0.570	0.455	JPN	Match 12	0.250		0.231	0.143	0.375	0.000	0.500	0.480	0.667			
Match 35		0.500	0.391	0.429	0.667	0.500	0.500	0.450	0.393	Match 28	0.500	0.429		0.455	0.313	1.000	0.500	0.560	0.486				
Match 49		0.000	0.161	0.154	0.500	0.500	0.500	0.370	0.333	Match 43	0.200	0.300		0.176	0.214	0.500	0.500	0.400	0.357				
C		ARG	Match 5	0.667	0.409	0.500	0.333	1.000	0.000	0.490	0.531	G		FRA	Match 13	0.500	0.563	0.429	0.800	0.000	0.000	0.510	0.471
		Match 21	1.000	0.733	0.900	0.429	0.500	0.500	0.580	0.606	Match 29			0.500	0.750	0.667	0.750	0.000	0.500	0.520	0.333		
		Match 37	0.500	0.526	0.500	0.588	0.000	0.000	0.470	0.568	Match 45			1.000	0.680	0.818	0.900	0.000	0.500	0.560	0.636		
	Match 50	0.667	0.478	0.625	0.545	0.500	0.500	0.510	0.568	Match 56	0.750		0.563	0.714	0.385	0.500	0.500	0.390	0.563				
	Match 57	0.500	0.545	0.500	0.600	0.333	0.000	0.580	0.434	Match 60	1.000		0.563	0.833	0.583	0.333	0.500	0.450	0.553				
	CIV	Match 5	0.333	0.591	0.500	0.667	0.000	1.000	0.510	0.469	Match 62		1.000	0.294	0.444	0.273	0.000	0.500	0.410	0.593			
	Match 22	0.333	0.640	0.529	0.727	0.000	1.000	0.490	0.605	Match 64	0.500		0.722	0.625	0.583	0.000	0.500	0.450	0.385				
	Match 38	0.600	0.769	0.769	0.900	0.000	0.500	0.680	0.586	SUI	Match 13		0.500	0.438	0.571	0.200	1.000	1.000	0.490	0.529			
	SCG	Match 6	0.000	0.478	0.400	0.600	0.667	1.000	0.390	0.611	Match 30		1.000	0.600	0.563	0.667	0.500	0.500	0.500	0.548			
	Match 21	0.000	0.267	0.100	0.571	0.500	0.500	0.420	0.394	Match 46	1.000		0.444	0.429	0.571	1.000	0.500	0.460	0.731				
	Match 38	0.400	0.231	0.231	0.100	1.000	0.500	0.320	0.414	Match 54	0.500		0.545	0.750	0.455	0.750	0.500	0.550	0.463				
	NED	Match 6	1.000	0.522	0.600	0.400	0.333	0.000	0.610	0.389	KOR		Match 14	0.667	0.640	0.667	0.429	0.667	0.000	0.640	0.531		
	Match 22	0.667	0.360	0.471	0.273	1.000	0.000	0.510	0.395	Match 29	0.500		0.250	0.333	0.250	1.000	0.500	0.480	0.667				
	Match 37	0.500	0.474	0.500	0.412	1.000	1.000	0.530	0.432	Match 46	0.000		0.556	0.571	0.429	0.000	0.500	0.540	0.269				
	Match 52	0.000	0.667	0.600	0.625	0.667	0.500	0.620	0.400	TOG	Match 14		0.333	0.360	0.333	0.571	0.333	1.000	0.360	0.469			
	MEX	Match 7	0.750	0.500	0.444	0.545	1.000	0.500	0.530	0.419	Match 30		0.000	0.400	0.438	0.333	0.500	0.500	0.500	0.452			
	Match 23	0.500	0.619	0.889	0.545	0.500	0.500	0.540	0.556	Match 45	0.000		0.320	0.182	0.100	1.000	0.500	0.440	0.364				
	Match 39	0.333	0.560	0.545	0.556	1.000	0.500	0.500	0.659	ESP	Match 15		1.000	0.792	0.833	0.875	1.000	0.500	0.540	0.560			
Match 50	0.333	0.522	0.375	0.455	0.500	0.500	0.490	0.432	Match 31	0.750	0.857	0.769	0.923	0.667	0.500	0.660	0.733						
D	IRN	Match 7	0.250	0.500	0.556	0.455	0.000	0.500	0.470	0.581	Match 47	1.000	0.731	0.765	0.714	0.000	0.500	0.590	0.488				
	Match 24	0.000	0.217	0.091	0.071	0.000	0.500	0.370	0.500	Match 56	0.250	0.438	0.286	0.615	0.500	0.500	0.610	0.438					
	Match 40	0.500	0.545	0.650	0.333	0.500	0.500	0.550	0.575	UKR	Match 15	0.000	0.208	0.167	0.125	0.000	0.500	0.460	0.440				
	ANG	Match 8	0.000	0.407	0.273	0.286	0.000	0.500	0.420	0.404	Match 32	1.000	0.760	1.000	0.750	0.500	0.500	0.500	0.489				
	Match 23	0.500	0.381	0.111	0.455	0.500	0.500	0.460	0.444	Match 48	1.000	0.500	0.667	0.500	0.333	0.500	0.470	0.575					
	Match 40	0.500	0.455	0.350	0.667	0.500	0.500	0.450	0.425	Match 54	0.500	0.455	0.250	0.545	0.250	0.500	0.450	0.537					
	POR	Match																					

表 4-5. 資料轉換後，把 32 隊各自資料加起來後的資料。

Team		$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$	$x_8$
	MP	GF	S	SG	CK	DFKG	IDFKG	BP	FS
GER(德國)	7.000	4.833	4.521	4.276	3.072	3.567	4.500	3.550	3.723
ECU(厄瓜多爾)	4.000	2.000	1.974	1.913	1.880	2.500	1.500	2.020	1.758
POL(波蘭)	3.000	0.667	1.104	1.189	1.932	1.400	2.000	1.490	1.553
CRC(哥斯大黎加)	3.000	0.667	1.167	0.947	1.071	0.600	1.500	1.350	1.553
ENG(英國)	5.000	3.500	2.806	2.892	2.907	3.500	2.500	2.630	2.647
SWE(瑞典)	4.000	2.000	1.818	2.102	2.722	2.500	2.000	1.990	1.769
PAR(巴拉圭)	3.000	1.500	1.475	1.335	1.351	2.000	1.500	1.430	1.610
TRI(千里達)	3.000	0.500	0.843	0.705	0.536	0.500	1.500	1.250	1.293
ARG(阿根廷)	5.000	3.333	2.693	3.025	2.496	2.333	1.000	2.630	2.707
NED(荷蘭)	4.000	2.167	2.022	2.171	1.710	3.000	1.500	2.270	1.616
CIV(象牙海岸)	3.000	1.267	2.000	1.799	2.294	0.000	2.500	1.680	1.660
SCG(塞爾維亞)	3.000	0.400	0.976	0.731	1.271	2.167	2.000	1.130	1.419
POR(葡萄牙)	7.000	4.500	4.064	4.354	4.367	3.933	3.000	3.820	3.605
MEX(墨西哥)	4.000	1.917	2.201	2.254	2.101	3.000	2.000	2.060	2.065
ANG(安哥拉)	3.000	1.000	1.243	0.734	1.407	1.000	1.500	1.330	1.274
IRN(伊朗)	3.000	0.750	1.263	1.297	0.859	0.500	1.500	1.390	1.656
ITA(義大利)	7.000	6.500	3.505	4.502	3.922	4.500	4.000	3.470	4.204
GHA(加納)	4.000	1.667	2.209	1.885	1.582	1.500	1.500	2.030	1.451
CZE(捷克)	3.000	1.000	1.477	1.738	1.620	1.500	2.000	1.430	1.514
USA(美國)	3.000	0.333	1.257	0.595	1.364	1.500	1.000	1.530	1.587
BRA(巴西)	5.000	3.800	2.641	2.845	2.684	2.667	2.500	2.670	2.919
AUS(澳大利亞)	4.000	1.250	2.257	2.357	2.181	1.500	2.000	2.130	1.463
CRO(克羅地亞)	3.000	1.000	1.381	1.179	1.579	2.000	1.500	1.380	1.548
JPN(日本)	3.000	0.950	0.959	0.774	0.902	1.500	1.500	1.440	1.510
SUI(瑞士)	4.000	3.000	2.027	2.313	1.893	3.250	2.500	2.000	2.272
FRA(法國)	7.000	5.250	4.134	4.531	4.274	0.833	3.000	3.290	3.533
KOR(韓國)	3.000	1.167	1.446	1.571	1.107	1.667	1.000	1.660	1.467
TOG(多哥)	3.000	0.333	1.080	0.953	1.005	1.833	2.000	1.300	1.284
ESP(西班牙)	4.000	3.000	2.817	2.653	3.128	2.167	2.000	2.400	2.219
UKR(烏克蘭)	5.000	2.500	2.488	2.583	2.671	1.583	2.500	2.470	2.406
TUN(突尼西亞)	3.000	0.750	0.959	0.850	1.006	1.000	1.000	1.360	1.120
KSA(沙特阿拉伯)	3.000	0.500	1.193	0.950	1.107	2.500	2.000	1.420	1.595

GF:Goals For S:Shots SG:Shots On Goal CK:Corner Kicks DFKG:Direct Free Kicks to Goal  
IDFKG:Indirect Free Kicks to Goal BP: Ball Possession FS:Fouls Committed

表 4-6. 32 隊各隊的單場平均輸入資料。

Team		$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$	$x_8$
	MP	GF	S	SG	CK	DFKG	IDFKG	BP	FS
GER(德國)	0.690	0.646	0.611	0.439	0.510	0.643	0.507	0.532	0.690
ECU(厄瓜多爾)	0.500	0.494	0.478	0.470	0.625	0.375	0.505	0.439	0.500
POL(波蘭)	0.222	0.368	0.396	0.644	0.467	0.667	0.497	0.518	0.222
CRC(哥斯大黎加)	0.222	0.389	0.316	0.357	0.200	0.500	0.450	0.518	0.222
ENG(英國)	0.700	0.561	0.578	0.581	0.700	0.500	0.526	0.529	0.700
SWE(瑞典)	0.500	0.454	0.525	0.681	0.625	0.500	0.498	0.442	0.500
PAR(巴拉圭)	0.500	0.492	0.445	0.450	0.667	0.500	0.477	0.537	0.500
TRI(千里達)	0.167	0.281	0.235	0.179	0.167	0.500	0.417	0.431	0.167
ARG(阿根廷)	0.667	0.538	0.605	0.499	0.467	0.200	0.526	0.541	0.667
NED(荷蘭)	0.542	0.506	0.543	0.427	0.750	0.375	0.568	0.404	0.542
CIV(象牙海岸)	0.422	0.667	0.600	0.765	0.000	0.833	0.560	0.553	0.422
SCG(塞爾維亞)	0.133	0.325	0.244	0.424	0.722	0.667	0.377	0.473	0.133
POR(葡萄牙)	0.643	0.581	0.622	0.624	0.562	0.429	0.546	0.515	0.643
MEX(墨西哥)	0.479	0.550	0.563	0.525	0.750	0.500	0.515	0.516	0.479
ANG(安哥拉)	0.333	0.414	0.245	0.469	0.333	0.500	0.443	0.425	0.333
IRN(伊朗)	0.250	0.421	0.432	0.286	0.167	0.500	0.463	0.552	0.250
ITA(義大利)	0.929	0.501	0.643	0.560	0.643	0.571	0.496	0.601	0.929
GHA(加納)	0.417	0.552	0.471	0.396	0.375	0.375	0.508	0.363	0.417
CZE(捷克)	0.333	0.492	0.579	0.540	0.500	0.667	0.477	0.505	0.333
USA(美國)	0.111	0.419	0.198	0.455	0.500	0.333	0.510	0.529	0.111
BRA(巴西)	0.760	0.528	0.569	0.537	0.533	0.500	0.534	0.584	0.760
AUS(澳大利亞)	0.313	0.564	0.589	0.545	0.375	0.500	0.533	0.366	0.313
CRO(克羅地亞)	0.333	0.460	0.393	0.526	0.667	0.500	0.460	0.516	0.333
JPN(日本)	0.317	0.320	0.258	0.301	0.500	0.500	0.480	0.503	0.317
SUI(瑞士)	0.750	0.507	0.578	0.473	0.813	0.625	0.500	0.568	0.750
FRA(法國)	0.750	0.591	0.647	0.611	0.119	0.429	0.470	0.505	0.750
KOR(韓國)	0.389	0.482	0.524	0.369	0.556	0.333	0.553	0.489	0.389
TOG(多哥)	0.111	0.360	0.318	0.335	0.611	0.667	0.433	0.428	0.111
ESP(西班牙)	0.750	0.704	0.663	0.782	0.542	0.500	0.600	0.555	0.750
UKR(烏克蘭)	0.500	0.498	0.517	0.534	0.317	0.500	0.494	0.481	0.500
TUN(突尼西亞)	0.250	0.320	0.283	0.335	0.333	0.333	0.453	0.373	0.250
KSA(沙特阿拉伯)	0.167	0.398	0.317	0.369	0.833	0.667	0.473	0.532	0.167

GF:Goals For S:Shots SG:Shots On Goal CK:Corner Kicks DFKG:Direct Free Kicks to Goal  
IDFKG:Indirect Free Kicks to Goal BP: Ball Possession FS:Fouls Committed



## 4.2 比賽結果的分類原則

爲了比較各種分類法分類結果的正確率，我們根據賽制的晉級與淘汰的特性，將比賽結果分成 4 類，如圖 4-1，第 A 類爲 1~4 名共 4 隊，第 B 類爲 5~8 名共 4 隊，第 C 類爲 9~16 名共 8 隊，第 D 類爲 17~32 名共 16 隊，因此由官方網站提供的比賽結果 [1] 轉換得到的 2006 年世界盃足球賽「比賽結果」分類對照表，如表 4-7 所示。

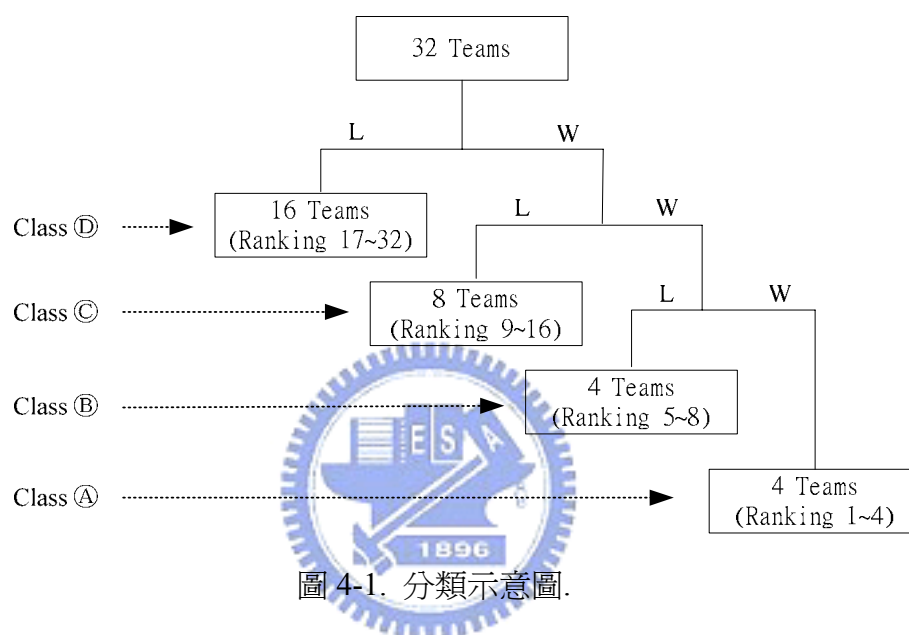


表 4-7. 2006 年世界盃足球賽「比賽結果」分類對照表.

賽後類別	名次	隊伍
Class A	1~4	義大利A、法國A、德國A、葡萄牙A
Class B	5~8	阿根廷B、英國B、巴西B、烏克蘭B
Class C	8~16	墨西哥C、瑞典C、瑞士C、西班牙C、澳大利亞C、加納C、荷蘭C、厄瓜多爾C
Class D	17~32	波蘭D、巴拉圭D、哥斯大黎加D、克羅地亞D、捷克D、安哥拉D、沙特阿拉伯D、象牙海岸D、伊朗D、美國D、塞爾維亞D、突尼西亞D、日本D、韓國D、多哥D、千里達D

### 4.3 分類正確率計算

我們將非監督式分類法的「分類結果」與世界盃的「比賽結果」作成如表 4-8 所示的分類結果比對表，n1~n16 代表符合的隊數，例如實驗分類後的 Class C1 這類的成員總共有 n1+n2+n3+n4 隊，其中有 n1 隊在比賽結束後的賽後類別是屬於 Class (A)，依此類推。

表 4-8. 分類結果比對表.

比賽結果 兩者符合 實驗結果 的隊數		賽後類別				正確隊數	正確率(%)
		Class (A)	Class (B)	Class (C)	Class (D)		
分類類別	Class C1	n1	n2	n3	n4		
	Class C2	n5	n6	n7	n8		
	Class C3	n9	n10	n11	n12		
	Class C4	n13	n14	n15	n16		

根據表 4-8 的比對表，我們在 4 個「分類類別 (Class C1~ Class C4)」對應的每一行中各取一個數後，把 4 個取出來的隊數相加成爲正確隊數，選取過程中若已被選取過隊數所對應的那一列的隊數即不能再被選取，因此列出下列 24 種 (S1 ~ S24) 可能的選取組合，來計算選出最高的值成爲該分類法的正確隊數。

$$\begin{aligned}
 S1 &= n1 + n6 + n11 + n16, & S2 &= n1 + n6 + n12 + n15, & S3 &= n1 + n7 + n10 + n16, \\
 S4 &= n1 + n7 + n12 + n14, & S5 &= n1 + n8 + n10 + n15, & S6 &= n1 + n8 + n11 + n14, \\
 S7 &= n2 + n5 + n11 + n16, & S8 &= n2 + n5 + n12 + n15, & S9 &= n2 + n7 + n9 + n16, \\
 S10 &= n2 + n7 + n12 + n13, & S11 &= n2 + n8 + n9 + n15, & S12 &= n2 + n8 + n11 + n13, \\
 S13 &= n3 + n5 + n10 + n16, & S14 &= n3 + n5 + n12 + n14, & S15 &= n3 + n6 + n9 + n16, \\
 S16 &= n3 + n6 + n12 + n13, & S17 &= n3 + n8 + n9 + n14, & S18 &= n3 + n8 + n10 + n13, \\
 S19 &= n4 + n5 + n10 + n15, & S20 &= n4 + n5 + n11 + n14, & S21 &= n4 + n6 + n9 + n15, \\
 S22 &= n4 + n6 + n11 + n13, & S23 &= n4 + n7 + n9 + n14, & S24 &= n4 + n7 + n10 + n13
 \end{aligned}$$

正確率計算公式爲：

$$\text{分類正確率(\%)} = \frac{\text{正確隊數}}{\text{總隊數}} \times 100\% \quad (4-1)$$

其中，總隊數=32 隊。

例如：經由 K-Means 演算法分類的「分類結果表 (表 4-9)」與「比賽結果表 (表 4-7)」比對後的「分類結果比對表」如 (表 4-10) 所示，經計算後可以得到 K-Means 演算法分類結果的正確隊數為 23 隊，換算成正確率為 71.875 %。

表4-9. K-Means 演算法分類結果表.

分類類別	K-Means 分類結果
Class C1	德國 <sup>Ⓐ</sup> 、義大利 <sup>Ⓐ</sup> 、葡萄牙 <sup>Ⓐ</sup> 、英國 <sup>Ⓑ</sup> 、巴西 <sup>Ⓑ</sup> 、西班牙 <sup>Ⓒ</sup> 、瑞士 <sup>Ⓒ</sup>
Class C2	法國 <sup>Ⓐ</sup> 、阿根廷 <sup>Ⓑ</sup> 、烏克蘭 <sup>Ⓑ</sup> 、象牙海岸 <sup>Ⓓ</sup>
Class C3	加納 <sup>Ⓒ</sup> 、澳大利亞 <sup>Ⓒ</sup> 、荷蘭 <sup>Ⓒ</sup> 、厄瓜多爾 <sup>Ⓒ</sup> 、瑞典 <sup>Ⓒ</sup> 、墨西哥 <sup>Ⓒ</sup> 、巴拉圭 <sup>Ⓓ</sup> 、捷克 <sup>Ⓓ</sup> 、韓國 <sup>Ⓓ</sup>
Class C4	突尼西亞 <sup>Ⓓ</sup> 、塞爾維亞 <sup>Ⓓ</sup> 、日本 <sup>Ⓓ</sup> 、多哥 <sup>Ⓓ</sup> 、美國 <sup>Ⓓ</sup> 、波蘭 <sup>Ⓓ</sup> 、克羅地亞 <sup>Ⓓ</sup> 、沙特阿拉伯 <sup>Ⓓ</sup> 、安哥拉 <sup>Ⓓ</sup> 、伊朗 <sup>Ⓓ</sup> 、千里達 <sup>Ⓓ</sup> 、哥斯大黎加 <sup>Ⓓ</sup>

表 4-10. K-Means 演算法分類結果比對表.

實驗結果	比賽結果 兩者符合 的隊數	賽後類別				正確隊數	正確率(%)
		Class <sup>Ⓐ</sup>	Class <sup>Ⓑ</sup>	Class <sup>Ⓒ</sup>	Class <sup>Ⓓ</sup>		
分類類別	Class C1	3	2	2	0	23	71.875
	Class C2	1	2	0	1		
	Class C3	0	0	6	3		
	Class C4	0	0	0	12		

## 5. 實驗結果

本論文的分類演算法實驗採用 MATLAB 7.0 版中的 Toolbox 提供的函數及實作而來，而實驗處理的過程，如：資料的擷取、資料的轉換、各種非監督式分類正確率的計算、分類的實驗結果、及分類個數的有效性驗證，皆利用 MATLAB 實作程式完成。各種非監督式分類的實驗過程及實驗結果，以及有效性驗證的實驗過程及實驗結果，分別如後小節所述。

### 5.1 非監督式分類法的實驗結果

#### (1) K-Means 分類結果

由於 K-Means 分類法不同的初始中心點及樣本出現順序會造成分類結果的不同，因此我們隨機選取初始中心點並將樣本隨機輸入重複實驗 10,000 次，將 10,000 次的每次正確隊數記下後，再求 10,000 次的平均值即為「平均的正確隊數」，同時把它換算成百分比即為「平均的正確率 (%)」，此外實驗過程中會同時紀錄下 10,000 次中曾經出現過的「最高正確隊數」，並把它換算成百分比即為「最高正確率 (%)」，另外同時會紀錄下最高正確隊數中其中一次的分類結果。

圖 5-1 為 10,000 次 K-Means 分類的正確隊數統計柱狀圖，其中最低的正确隊數為 11 隊，出現過 7 次，而最高正確隊數為 23 隊，出現過 48 次，而出現過最多次數的正確隊數為 14 隊，有 2,288 次。

最高正確隊數 23 隊 48 次中，其中 1 次的分類結果列於表 5-1，該次分類結果的各類中心列於表 5-2，此外，K-Means 分類的正確隊數及正確率列於表 5-3。

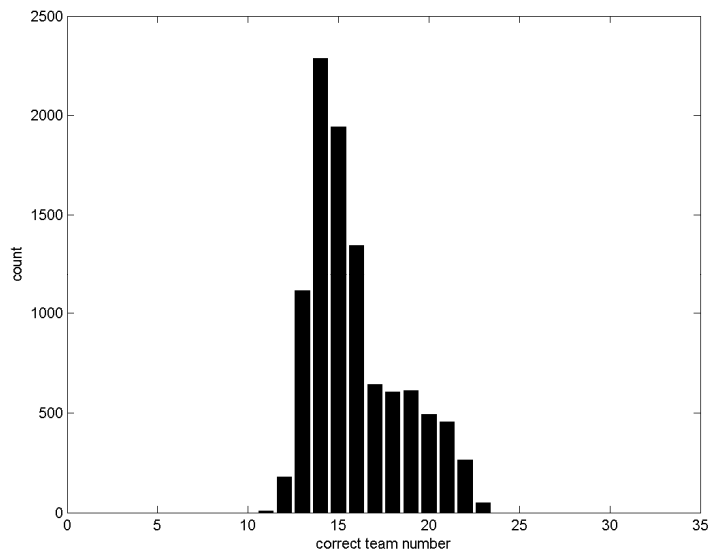


圖 5-1. 10,000 次 K-Means 分類的正確隊數統計柱狀圖.

表 5-1. K-Means 分類結果表 (48 次最高正確隊數中，其中 1 次的的分類結果).

類別	K-Means 分類結果
Class C1	德國 <sup>Ⓐ</sup> 、義大利 <sup>Ⓐ</sup> 、葡萄牙 <sup>Ⓐ</sup> 、英國 <sup>Ⓑ</sup> 、巴西 <sup>Ⓑ</sup> 、西班牙 <sup>Ⓒ</sup> 、瑞士 <sup>Ⓒ</sup>
Class C2	法國 <sup>Ⓐ</sup> 、阿根廷 <sup>Ⓑ</sup> 、烏克蘭 <sup>Ⓑ</sup> 、象牙海岸 <sup>Ⓓ</sup>
Class C3	加納 <sup>Ⓒ</sup> 、澳大利亞 <sup>Ⓒ</sup> 、荷蘭 <sup>Ⓒ</sup> 、厄瓜多爾 <sup>Ⓒ</sup> 、瑞典 <sup>Ⓒ</sup> 、墨西哥 <sup>Ⓒ</sup> 、巴拉圭 <sup>Ⓓ</sup> 、捷克 <sup>Ⓓ</sup> 、韓國 <sup>Ⓓ</sup>
Class C4	突尼西亞 <sup>Ⓓ</sup> 、塞爾維亞 <sup>Ⓓ</sup> 、日本 <sup>Ⓓ</sup> 、多哥 <sup>Ⓓ</sup> 、美國 <sup>Ⓓ</sup> 、波蘭 <sup>Ⓓ</sup> 、克羅地亞 <sup>Ⓓ</sup> 、沙特阿拉伯 <sup>Ⓓ</sup> 、安哥拉 <sup>Ⓓ</sup> 、伊朗 <sup>Ⓓ</sup> 、千里達 <sup>Ⓓ</sup> 、哥斯大黎加 <sup>Ⓓ</sup>

表 5-2. 表 5-1 該次 K-Means 最高正確隊數分類結果的各類中心.

類別 \ 中心	GF	S	SOG	CK	DFKG	IFKG	BP	FS
Class C1	0.21806	0.37292	0.30283	0.13595	0.22963	0.45093	0.45472	0.49168
Class C2	0.44136	0.50955	0.52428	0.22347	0.32099	0.49074	0.51463	0.45424
Class C3	0.58472	0.57334	0.59211	0.22073	0.11667	0.24345	0.5125	0.52126
Class C4	0.74599	0.5754	0.60926	0.4148	0.46573	0.52245	0.5298	0.54198

表 5-3. K-Means 分類的正確隊數及正確率.

非監督式分類演算法	平均 正確隊數	平均正確率(%) (總隊數 32 隊)	最高的 正確隊數	最高的正確率 (%)
K-Means 分類法	15.9273	49.772 %	23	71.875 %

## (2) 合併式階層分類法實驗結果

首先我們將世界盃的輸入資料取歐式距離後，再利用 MATLAB Statistics Toolbox 中的 linkage 函數，分別採用 single link、complete link、及 average link，3 種不同連結的參數取得 hierarchical cluster tree 後，再以 dendrogram 函數把結果畫成如圖 5-2 (a) (b) (c)所示 3 種不同的合併式階層樹狀圖，從圖中設定一個分成 4 類的分類準位，然後把合併式階層分類法的分類結果列於表 5-4，同時計算合併式階層分類法的分類正確率，然後把正確率列於表 5-5，此實驗中 3 種不同連結法各自只需實驗一次，因為每次實驗的結果皆相同。

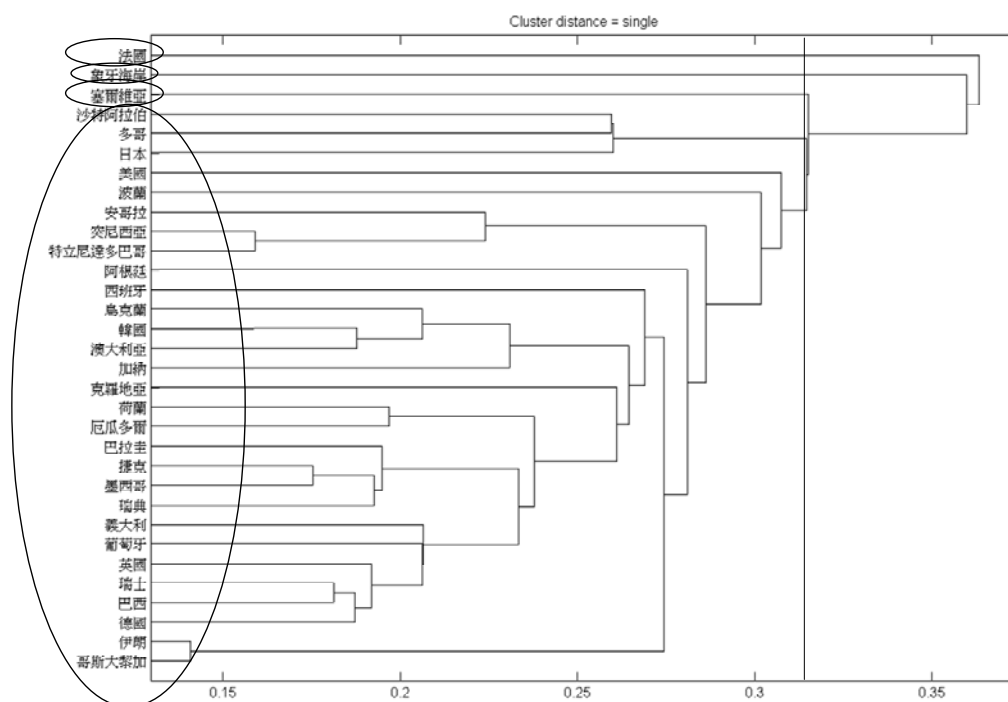


圖 5-2. (a) Single link 合併式階層分類結果樹狀圖.

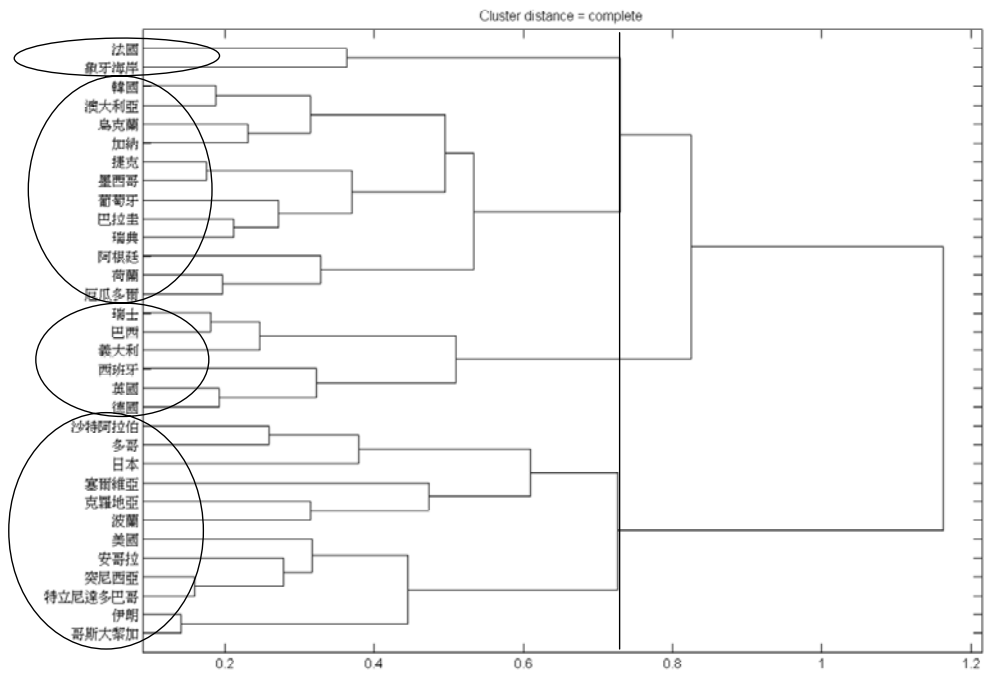


圖 5-2. (b) Complete link 合併式階層分類結果樹狀圖.

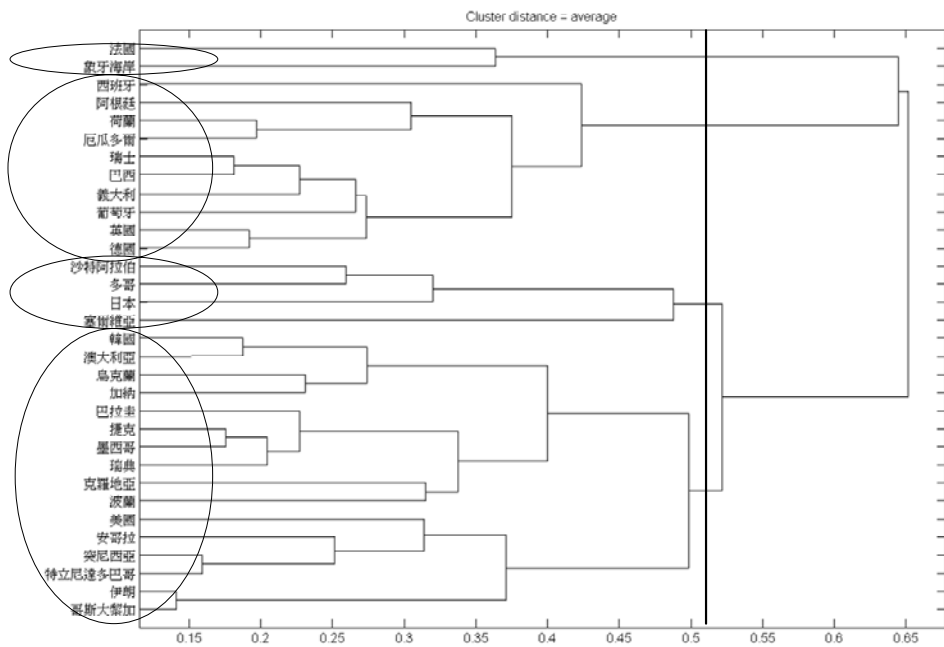


圖 5-2. (c) Average link 合併式階層分類結果樹狀圖.



表 5-4. 合併式階層分類結果表.

合併式階層法	類別	合併式階層分類結果
Single link	Class C1	法國 <sup>Ⓐ</sup>
	Class C2	象牙海岸 <sup>Ⓓ</sup>
	Class C3	塞爾維亞 <sup>Ⓓ</sup>
	Class C4	德國 <sup>Ⓐ</sup> 、義大利 <sup>Ⓐ</sup> 、葡萄牙 <sup>Ⓐ</sup> 、巴西 <sup>Ⓑ</sup> 、英國 <sup>Ⓑ</sup> 、烏克蘭 <sup>Ⓑ</sup> 、阿根廷 <sup>Ⓑ</sup> 西班牙 <sup>Ⓒ</sup> 、墨西哥 <sup>Ⓒ</sup> 、瑞典 <sup>Ⓒ</sup> 、荷蘭 <sup>Ⓒ</sup> 、加納 <sup>Ⓒ</sup> 、澳大利亞 <sup>Ⓒ</sup> 、瑞士 <sup>Ⓒ</sup> 厄瓜多爾 <sup>Ⓒ</sup> 、巴拉圭 <sup>Ⓓ</sup> 、哥斯大黎加 <sup>Ⓓ</sup> 、安哥拉 <sup>Ⓓ</sup> 、沙特阿拉伯 <sup>Ⓓ</sup> 、突尼西亞 <sup>Ⓓ</sup> 、 日本 <sup>Ⓓ</sup> 、多哥 <sup>Ⓓ</sup> 、捷克 <sup>Ⓓ</sup> 、波蘭 <sup>Ⓓ</sup> 、伊朗 <sup>Ⓓ</sup> 、克羅地亞 <sup>Ⓓ</sup> 、韓國 <sup>Ⓓ</sup> 、美國 <sup>Ⓓ</sup> 千里達 <sup>Ⓓ</sup>
Complete link	Class C1	法國 <sup>Ⓐ</sup> 、象牙海岸 <sup>Ⓓ</sup>
	Class C2	韓國 <sup>Ⓓ</sup> 、澳大利亞 <sup>Ⓒ</sup> 、烏克蘭 <sup>Ⓑ</sup> 、加納 <sup>Ⓒ</sup> 、捷克 <sup>Ⓓ</sup> 、墨西哥 <sup>Ⓒ</sup> 、葡萄牙 <sup>Ⓐ</sup> 、巴拉圭 <sup>Ⓓ</sup> 、瑞典 <sup>Ⓒ</sup> 、阿根廷 <sup>Ⓑ</sup> 、荷蘭 <sup>Ⓒ</sup> 、厄瓜多爾 <sup>Ⓒ</sup>
	Class C3	德國 <sup>Ⓐ</sup> 、義大利 <sup>Ⓐ</sup> 、英國 <sup>Ⓑ</sup> 、巴西 <sup>Ⓑ</sup> 、西班牙 <sup>Ⓒ</sup> 、瑞士 <sup>Ⓒ</sup>
	Class C4	沙特阿拉伯 <sup>Ⓓ</sup> 、塞爾維亞 <sup>Ⓓ</sup> 、日本 <sup>Ⓓ</sup> 、美國 <sup>Ⓓ</sup> 、突尼西亞 <sup>Ⓓ</sup> 、安哥拉 <sup>Ⓓ</sup> 千里達 <sup>Ⓓ</sup> 、伊朗 <sup>Ⓓ</sup> 、哥斯大黎加 <sup>Ⓓ</sup> 、多哥 <sup>Ⓓ</sup> 、克羅地亞 <sup>Ⓓ</sup> 、波蘭 <sup>Ⓓ</sup>
Average link	Class C1	法國 <sup>Ⓐ</sup> 、象牙海岸 <sup>Ⓓ</sup>
	Class C2	西班牙 <sup>Ⓒ</sup> 、阿根廷 <sup>Ⓑ</sup> 、荷蘭 <sup>Ⓒ</sup> 、厄瓜多爾 <sup>Ⓒ</sup> 、瑞士 <sup>Ⓒ</sup> 、巴西 <sup>Ⓑ</sup> 、義大利 <sup>Ⓐ</sup> 、葡萄牙 <sup>Ⓐ</sup> 、英國 <sup>Ⓑ</sup> 、德國 <sup>Ⓐ</sup>
	Class C3	沙特阿拉伯 <sup>Ⓓ</sup> 、多哥 <sup>Ⓓ</sup> 、日本 <sup>Ⓓ</sup> 、塞爾維亞 <sup>Ⓓ</sup>
	Class C4	韓國 <sup>Ⓓ</sup> 、澳大利亞 <sup>Ⓒ</sup> 、烏克蘭 <sup>Ⓑ</sup> 、加納 <sup>Ⓒ</sup> 、巴拉圭 <sup>Ⓓ</sup> 、捷克 <sup>Ⓓ</sup> 、墨西哥 <sup>Ⓒ</sup> 、瑞典 <sup>Ⓒ</sup> 、克羅地亞 <sup>Ⓓ</sup> 、波蘭 <sup>Ⓓ</sup> 、美國 <sup>Ⓓ</sup> 、安哥拉 <sup>Ⓓ</sup> 、突尼西亞 <sup>Ⓓ</sup> 、千里達 <sup>Ⓓ</sup> 、伊朗 <sup>Ⓓ</sup> 、哥斯大黎加 <sup>Ⓓ</sup>

表 5-5. 合併式階層分類的正確隊數及正確率.

非監督式分類演算法	正確隊數	正確率(%) (總隊數 32 隊)	最高的 正確隊數	最高的正確率 (%)
Single link 合併式階層分類法	15	46.87 %		
Complete link 合併式階層分類法	22	68.75 %		
Average link 合併式階層分類法	16	50 %		

### (3) Fuzzy C-means 分類結果

由於 Fuzzy C-Means 演算法不同的隨機初始值及樣本出現順序會有不同的結果，因此我們隨機選取樣本輸入的順序，並重複實驗 10,000 次 Fuzzy C-Means 的分類，實驗函數採用 MATLAB Fuzzy Toolbox 的 FCM 函數，記錄實驗結果的作法如同 K-Means，圖 5-3 為 10,000 次 Fuzzy C-Means 分類的正確隊數統計柱狀圖，該圖中最低的正确隊數為 14 隊，出現過 1,115 次，而最高正确隊數為 17 隊，出現過 7,912 次，而出現過最多次數的正确隊數為 17 隊，有 7,912 次，這顯示 Fuzzy C-Means 每次分類結果的正确率比 K-Means 的結果而言，相當穩定。在最高正确隊數的 7,912 次中，其中 1 次的分類結果列於表 5-6，Fuzzy C-Means 分類法的分類正确隊數及分類正确率列於表 5-7。

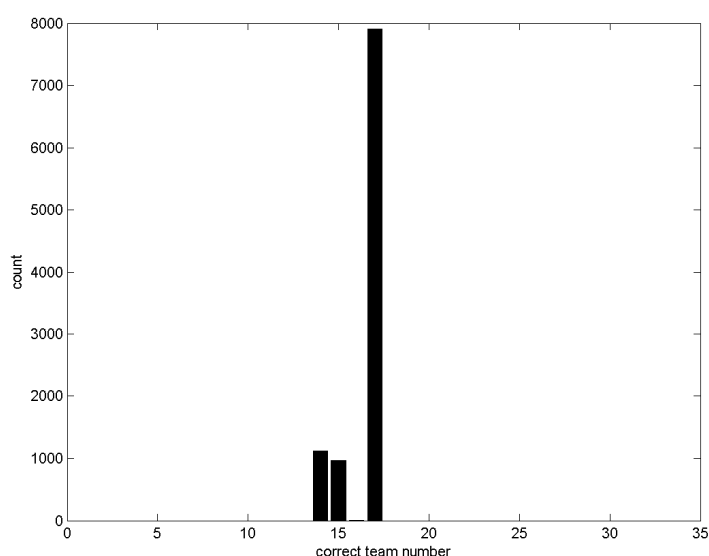


圖 5-3. 10,000 次 Fuzzy C-Means 分類的正确隊數統計柱狀圖.

表 5-6. Fuzzy C-Means 分類結果表 (最高正确隊數 7,912 次中，其中 1 次的分類結果).

類別	Fuzzy C-Means 分類結果
Class C1	德國 <sup>Ⓐ</sup> 、葡萄牙 <sup>Ⓐ</sup> 、義大利 <sup>Ⓐ</sup> 、英國 <sup>Ⓑ</sup> 、阿根廷 <sup>Ⓑ</sup> 、巴西 <sup>Ⓑ</sup> 、瑞士 <sup>Ⓒ</sup> 、西班牙 <sup>Ⓒ</sup>
Class C2	法國 <sup>Ⓐ</sup> 、烏克蘭 <sup>Ⓑ</sup> 、澳大利亞 <sup>Ⓒ</sup> 、加納 <sup>Ⓒ</sup> 、伊朗 <sup>Ⓓ</sup> 、象牙海岸 <sup>Ⓓ</sup> 、韓國 <sup>Ⓓ</sup>
Class C3	厄瓜多爾 <sup>Ⓒ</sup> 、墨西哥 <sup>Ⓒ</sup> 、瑞典 <sup>Ⓒ</sup> 、巴拉圭 <sup>Ⓓ</sup> 、荷蘭 <sup>Ⓒ</sup> 、捷克 <sup>Ⓓ</sup> 、克羅地亞 <sup>Ⓓ</sup> 、沙特阿拉伯 <sup>Ⓓ</sup>
Class C4	波蘭 <sup>Ⓓ</sup> 、哥斯大黎加 <sup>Ⓓ</sup> 、安哥拉 <sup>Ⓓ</sup> 、突尼西亞 <sup>Ⓓ</sup> 、千里達 <sup>Ⓓ</sup> 、塞爾維亞 <sup>Ⓓ</sup> 、美國 <sup>Ⓓ</sup> 、日本 <sup>Ⓓ</sup> 、多哥 <sup>Ⓓ</sup>

表 5-7. Fuzzy C-Means 分類的正確隊數及正確率。

非監督式分類演算法	平均 正確隊數	平均正確率(%) (總隊數 32 隊)	最高的 正確隊數	最高的正確率 (%)
Fuzzy C-Means 分類法	16.4713	51.472 %	17	53.125 %

#### (4) SOFM 分類結果

我們針對如圖 5-4 (a) (b)(c) 所示 3 種不同 output node topology 的 SOFM 進行實驗，output nodes 為四個，距離採 Euclidian 距離，epochs=100，勝利輸出神經元 weighting vector

調整時的 learning rate  $\eta_{j^*}(t) = \alpha(1 - \frac{t}{100})$ ,  $\alpha = 0.9$ ，而勝利神經元鄰近函數  $NE_{j^*}(t)$  內的神經

元 weighting vector 調整時之 learning rate  $\eta_j(t) = \frac{1}{2}\alpha(1 - \frac{t}{100})$ ,  $\alpha = 0.9$ ， $NE_{j^*}(t)$  的影響半徑隨

時間增加而變小，其影響半徑調整公式  $NE_{j^*}(t) = 1.00001 + (ND_{\max} - 1)(1 - \frac{t}{100})$ ，其中  $ND_{\max}$  為

兩兩輸出神經元之間距離最遠的距離，最終影響半徑的範圍  $NE_{j^*}(100) = 1$ 。每一種 topology

SOFM 各實驗 200 次，然後記錄下 200 次每次實驗的正確隊數之後，再取 200 次實驗的平均值即為「平均正確隊數」，同時把它換算成百分比成為「平均正確率」，此外實驗過程中會同時記錄 200 次實驗中，曾出現過的最高正確隊數成為「最高正確隊數」，並換算成百分比即為「最高正確率」，另外同時會紀錄最高正確隊數其中 1 次的分類結果。

圖 5-5 為  $\eta_{j^*}(t)$  與 t 的關係圖，圖 5-6 (a)(b)(c) 為 3 種不同 Topology SOFM 分類法 200 次實驗結果的正確隊數統計柱狀圖，其中：

- (1) 1D Topology SOFM 的最高正確隊數為 22 隊，最低的正确隊數為 12 隊，出現過 1 次。
- (2) 2D GridTop Topology SOFM 的最高正確隊數為 21 隊，出現過 20 次，最低的正确隊數為 14 隊，出現過 9 次。
- (3) 2D HexTop Topology SOFM 的最高正確隊數為 21 隊，出現過 18 次，最低的正确隊數為 12 隊，出現過 1 次。

3 種不同 Topology SOFM 最高正確隊數其中 1 次的分類結果，並列於表 5-8，而 3 種不同 SOFM 分類的正確隊數及正確率，並列於表 5-9。

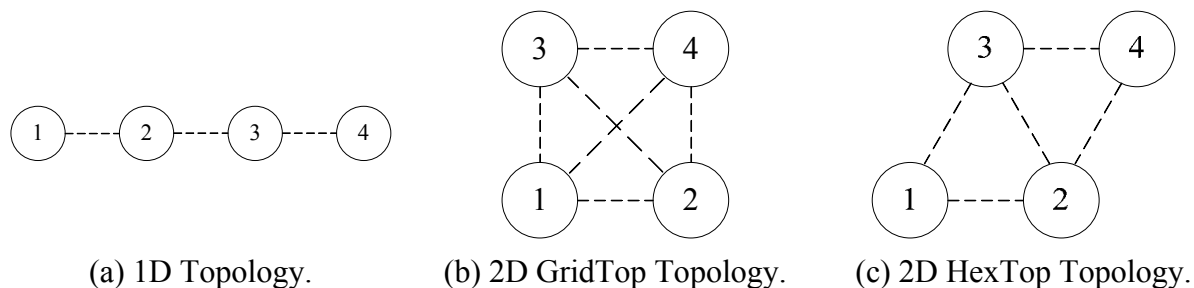


圖 5-4. 2006 年世界盃足球賽 SOFM 分類演算法分成 4 類的 output nodes topology.

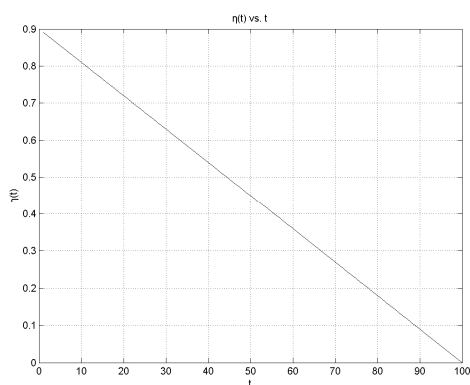


圖 5-5.  $\eta_j^*(t)$  與  $t$  的關係圖.

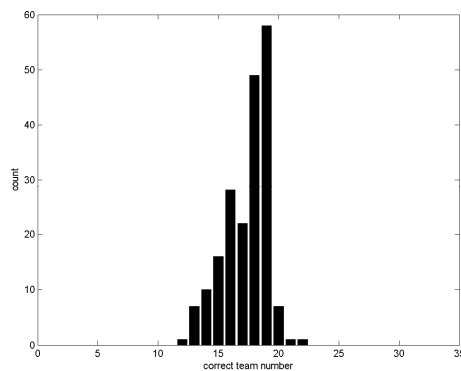


圖 5-6. (a) 200 次 1D Topology SOFM 之正確隊數統計柱狀圖.

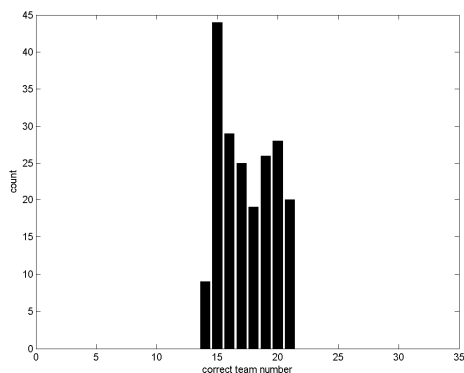


圖 5-6. (b) 200 次 2D GridTop Topology SOFM 之正確隊數統計柱狀圖.

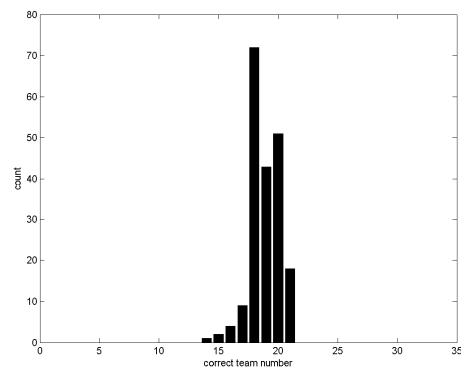


圖 5-6. (c) 200 次 2D HexTop Topology SOFM 之正確隊數統計柱狀圖.

表 5-8. SOFM 分類結果表 (最高正確隊數其中 1 次的分類結果).

SOFM	類別	SOFM 分類結果
1D SOFM	Class C1	德國(A)、義大利(A)、葡萄牙(A)、英國(B)、巴西(B)、西班牙(C)、荷蘭(C)、瑞士(C)
	Class C2	法國(A)、烏克蘭(B)、阿根廷(B)、象牙海岸(D)
	Class C3	澳大利亞(C)、捷克(D)、加納(C)、厄瓜多爾(C)、瑞典(C)、墨西哥(C)、巴拉圭(D)、韓國(D)
	Class C4	克羅地亞(D)、哥斯大黎加(D)、安哥拉(D)、沙特阿拉伯(D)、塞爾維亞(D)、突尼西亞(D)、日本(D)、多哥(D)、千里達(D)、伊朗(D)、波蘭(D)、美國(D)
2D GridTop SOFM	Class C1	義大利(A)、德國(A)、葡萄牙(A)、法國(A)、英國(B)、巴西(B)、阿根廷(B)、西班牙(C)、瑞士(C)
	Class C2	烏克蘭(B)、澳大利亞(C)、象牙海岸(D)、韓國(D)
	Class C3	厄瓜多爾(C)、荷蘭(C)、墨西哥(C)、瑞典(C)、巴拉圭(D)、捷克(D)
	Class C4	沙特阿拉伯(C)、美國(D)、安哥拉(D)、多哥(D)、波蘭(D)、克羅地亞(D)、千里達(D)、伊朗(D)、塞爾維亞(D)、哥斯大黎加(D)、突尼西亞(D)、日本(D)、加納(C)
2D HexTop SOFM	Class C1	德國(A)、法國(A)、義大利(A)、葡萄牙(A)、英國(B)、阿根廷(B)、巴西(B)、西班牙(C)、瑞士(C)
	Class C2	烏克蘭(B)、加納(C)、澳大利亞(C)、象牙海岸(D)、捷克(D)
	Class C3	荷蘭(C)、墨西哥(C)、厄瓜多爾(C)、瑞典(C)、巴拉圭(D)、韓國(D)
	Class C4	美國(D)、千里達(D)、伊朗(D)、塞爾維亞(D)、哥斯大黎加(D)、日本(D)、安哥拉(D)、克羅地亞(D)、波蘭(D)、多哥(D)、突尼西亞(D)、沙特阿拉伯(D)

表 5-9. SOFM 分類正確隊數及正確率.

非監督式分類演算法	平均 正確隊數	平均正確率(%) (總隊數 32 隊)	最高的 正確隊數	最高的正確率 (%)
1D Topology SOFM	17.36	54.25 %	22	68.75 %
2D GridTop Topology SOFM	17.45	54.54 %	21	65.625 %
2D HexTop Topology SOFM	18.86	58.93 %	21	65.625 %

### (5) 非監督式分類法的平均正確率及最高正確率總比較

2006 年世界盃足球賽統計數據經由各種非監督式分類法把 32 隊分類的分類正確率結果列於表 5-10 的比較表，從這個比較表中，我們看出各種非監督式分類法的分類平均正確率約介於 46.87 % ~ 68.75 % 之間，其中合併式階層演算法的完全連結法擁有最高的平均正確率 68.75 %，而合併式階層分類法中的單一連結法擁有最低的平均準確率 46.87%，不到 50%，其最高的正確率則出現在 K-Means 分類法可達 71.875 %。

此外我們進一步比較圖 5-1 的 K-Means 及圖 5-3 的 Fuzzy C-Means 分類結果之正確隊數的分佈圖，由圖中可以看出 K-Means 的正確隊數分佈範圍較為分散，從 11~23 隊皆有，而 Fuzzy C-Means 的正確隊數分佈範圍則比較集中，主要落在 17 隊。

表 5-10. 非監督式分類法實驗結果之正確率比較表.

非監督式分類演算法		平均 正確隊數	平均正確率(%) (總隊數 32 隊)	最高的 正確隊數	最高的正確率 (%)
K-Means		15.9273	49.772 %	23	71.875 %
Fuzzy C-Means		16.4713	51.472 %	17	53.125 %
合併階層演算法	Single link	15	46.87 %		
	Complete link	22	68.75 %		
	Average link	16	50 %		
SOFM	1D TOP	17.36	54.25 %	22	68.75 %
	2D GRIDTOP	17.45	54.54 %	21	65.625 %
	2D HEXTOP	18.86	58.93 %	21	65.625 %

## 5.2 2006 年世界盃足球賽統計數據分類個數有效性的驗證結果

我們利用 3 種分類有效性驗證演算法來驗證 2006 年世界盃足球賽的統計數據，其分別為：(1) BSC 有效性驗證，(2) MBSC 有效性驗證，(3) PFS 有效性驗證。

### 利用以 Sequential Cluster 為基礎的有效性驗證演算法

由 2006 年世界盃足球賽的資料中，我們計算找到最接近的兩個樣本之間的最小歐式距離  $a=0.0704$ ，及最遠離的兩個樣本之間的最大歐式距離為  $b=1.1635$ ，並把 step size 設為  $c = (b-a)/200=0.0055$ ， $s=200$ ，每一個門檻設定值皆利用 BSC 及 MBSC 兩種有效性驗證演算法各測試 1,000 次後找出最常出現的 K，然後將其門檻值與 K 值的關係圖畫出，利用 BSC 及 MBSC 的有效性驗證結果如圖 5-7 及圖 5-8 所示。

由圖 5-7 的 BSC 有效性驗證的結果圖中，我們可以清楚地看到在  $K=2$  時有明顯的轉折變化，其意謂著 2006 年世界盃足球賽球隊的統計資料具有群聚的特性，且其最佳分類個數 K 值等於 2。

同理，在圖 5-8 的 MBSC 有效性驗證的結果圖中指出利用 MBSC 驗證法的最佳分類個數 K 值等於 2，其次 K 值等於 3。



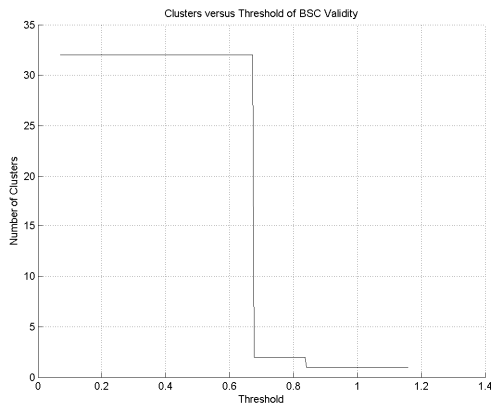
### 利用以 KM 為基礎的 PFS 有效性驗證演算法

因 PFS 演算法以 KM 為基礎，由於 KM 演算法不同的初始中心及樣本出現順序的不同會造成分類的結果不同，若只根據一次的 PFS 驗證結果做為最佳分類個數 K 值則不夠客觀，所以針對 2006 年世界盃足球賽的 32 隊，我們分別測試 K 值從 2~16，每一個 K 值反覆實驗 10,000 次後可得到 10,000 次該 K 值的 PFS 值，最後再取這 10,000 次 PFS 的平均值，每一次實驗的輸入都經過亂數選取中心點及隨機的樣本輸入順序。PFS 有效性驗證的結果如圖 5-9 所示，從該圖中的 PFS 曲線，我們可以找當到  $K=2$  時 PFS 具有最大值，其次為  $K=3$  時 PFS 具有次大的值，所以最佳的分類個數為 2 類，其次為 3 類。

綜合上述 3 種分類有效性驗證的結果，我們可以得到一個結論，2006 世界盃足球賽的資料若欲分類，它的分類個數以分成 2 類為最佳的選擇，其次為分成 3 類。



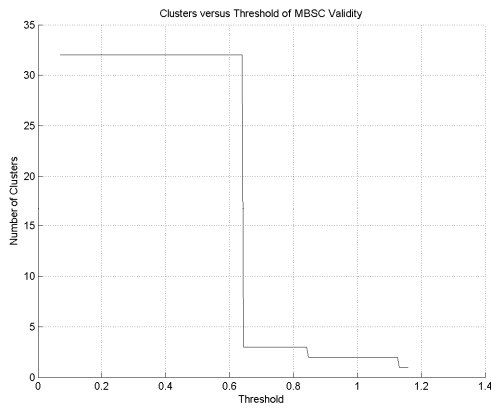
### (1) BSC 有效性驗證結果



$\theta$	<b>K</b>
$0.0704 \leq \theta < 0.6770$	32
$0.6770 \leq \theta < 0.8410$	2
$0.8410 \leq \theta < 1.1635$	1

圖 5-7. BSC 分類有效性結果圖.

### (2) MBSC 有效性驗證結果



$\theta$	<b>K</b>
$0.0704 \leq \theta < 0.6442$	32
$0.6442 \leq \theta < 0.8464$	3
$0.8464 \leq \theta < 1.1307$	2
$1.1307 \leq \theta < 1.1635$	1

圖 5-8. MBSC 分類有效性結果圖.

### (3) PFS 有效性驗證結果

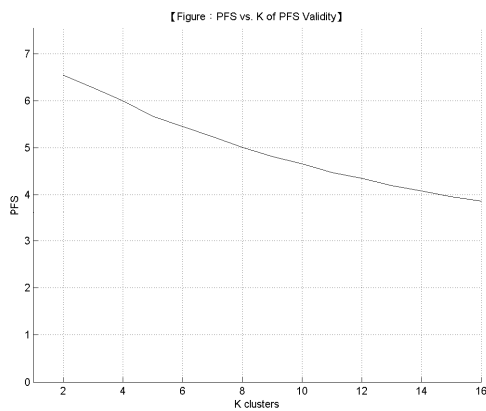


圖 5-9. PFS 分類有效性結果圖.

### 5.3 主分量分析法用於 2006 年世界盃足球賽統計數據之分類實驗結果

此節的實驗目的，主要是利用主分量分析法 (Principal Component Analysis) 來分析 2006 年世界盃統計數據後，將原始資料重新投影到新選取之主要成分的軸上，來降低資料的維度並減少雜訊對分類造成的影響，但同時仍保持資料在空間上的分佈特性。

首先我們利用 MATLAB Toolbox 中的 princomp 函數將原始資料經過主分量分析後，得到如表 5-11 所列，在 8 個主軸基底上的 8 個 feature 的 coefficient，及得到如表 5-12 所列，covariance matrix 在 8 個主軸基底上之 eigenvalues，total eigenvalues 的百分比，及 total eigenvalues 累積的百分比，。

表 5-12 第一欄為 eigenvalues，此欄可以看出第 1 組基底 PC1 上的 eigenvalues 最大，而第 8 組基底 PC8 上的 eigenvalues 最小，第二欄為每一基底上的 total eigenvalues 的百分比，第三欄為從第 1 組基底到第 8 組基底所累積的 total eigenvalues 的百分比。

圖 5-10 所示為經 PCA 分析的 2006 世界盃足球賽輸入資料，投影至 8 個主軸上 total eigenvalues 所累積的百分比直條圖。

表 5-11. 2006 世界盃足球賽分類輸入資料之 principal component and feature coefficient.

PC Feature Coe.	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
x1 coefficient 1	-0.6862	0.24846	-0.22759	0.60179	-0.11256	-0.1803	-0.08805	-0.022944
x2 coefficient 2	-0.26467	0.17435	-0.02405	-0.50601	-0.14844	-0.014474	-0.71074	0.3399
x3 coefficient 3	-0.4124	0.21962	-0.08353	-0.45086	0.2978	0.37181	0.58695	0.0022308
x4 coefficient 4	-0.37005	-0.12623	-0.027912	-0.28276	0.82199	-0.26159	0.14878	0.02609
x5 coefficient 5	-0.36366	-0.59469	0.68272	0.059278	-0.20265	0.024984	-0.0091462	0.051851
x6 coefficient 6	-0.069474	-0.69796	-0.68548	-0.072888	-0.15816	0.055908	-0.056709	-0.038374
x7 coefficient 7	-0.09924	0.052242	0.061806	-0.20664	-0.061694	-0.043302	-0.23951	-0.93694
x8 coefficient 8	-0.071751	-0.01291	-0.0072699	0.21693	0.362	0.8689	-0.24451	-0.041972

表 5-12. 2006 世界盃足球賽分類輸入資料之 covariance matrix 的 eigenvalues 及 total eigenvalues 百分比，及 total eigenvalues 累積的百分比。

PC	Eigenvalues of the covariance matrix	Percentage of the total eigenvalues	Accumulated percentage of the total eigenvalues
PC1	0.094576	55.955 %	<b>55.955 %</b>
PC2	0.039184	23.183 %	<b>79.138 %</b>
PC3	0.017196	10.174 %	<b>89.311 %</b>
PC4	0.0076765	4.5417 %	<b>93.853 %</b>
PC5	0.0056189	3.3243 %	97.177 %
PC6	0.0023785	1.4072 %	98.585 %
PC7	0.0017106	1.0121 %	99.597 %
PC8	0.00068188	0.40343 %	100 %

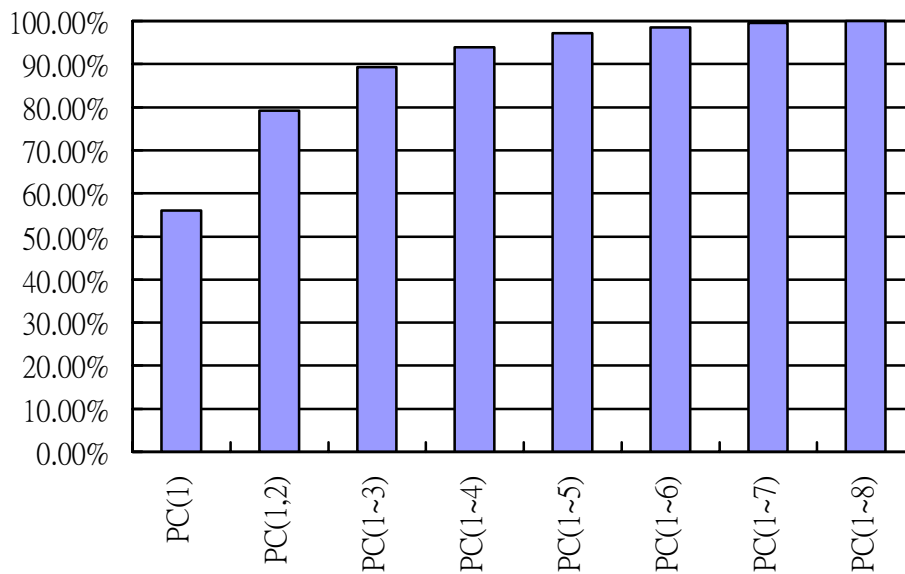


圖 5-10. 2006 世界盃足球賽輸入資料經 PCA 分析後投影至 8 個主軸上，total eigenvalues 所累積的百分比直條圖。

此外，我們分別從取前 2 組基底 PC(1,2) 做為 features，取前 3 組基底 PC(1~3) 做為 features，依此類推，一直到取前 8 組基底 PC(1~8) 做為 features，共 8 種取法，每 1 種取法皆採用 K-Means 及 Fuzzy C-Means 來將資料重新分成 4 類，每 1 種取法各實驗 10,000 次後計算出平均正確隊數，再與未做 PCA 處理之分類結果做比較，經 PCA 分析後，取不同 feature 個數之 K-Means 及 Fuzzy C-Means 的分類正確隊數比較表列於表 5-13，其分類正確隊數的曲線圖則畫於圖 5-11。

由表 5-12 第三欄及圖 5-10 的直條圖中，我們可以看出將 2006 年世界盃足球賽的輸入資料投影到前 4 個基底上的累積的 total eigenvalues 擁有超過 93.85 % 的貢獻度，且由表 5-13 及圖 5-11 的分類正確隊數比較結果，我們發現利用 PCA 處理後投影到前 4 個主軸基底上的分類正確隊數即可接近未經 PCA 處理過的分類正確隊數。

因此，根據這樣的觀察，我們決定先把輸入資料投影到前 4 個主軸基底上，然後再利用各種非監督式分類法把 2006 世界盃足球賽的統計數據重新分類，分類的個數依 5.2 節實驗的結果為分成 3 類及 2 類，每一種演算法只做一次，各種非監督式分類法的分類結果列於表 5-14 ~ 表 5-17，其分別為未經 PCA 處理及取 4 個 features 經過 PCA 處理後，重新分成 3 類及分成 2 類的分類結果。

表 5-13. 經 PCA 分析後取不同 feature 個數，K-Means 及 Fuzzy C-Means 分類結果的正確隊數比較表.

分類法 \ 投影基底	PC(1,2)	PC(1~3)	<b>PC(1~4)</b>	PC(1~5)	PC(1~6)	PC(1~7)	PC(1~8)	No PCA
<b>KM (average)</b>	15.4557	15.5572	<b>15.7325</b>	15.870	15.8713	15.9030	15.9287	15.9273
<b>FCM (average)</b>	16.3558	16.4487	<b>16.4624</b>	16.4483	16.4830	16.4660	16.4403	16.4713
KM (maximum)	23	24	<b>23</b>	23	23	23	23	23
FCM (maximum)	17	17	<b>17</b>	17	17	17	17	17

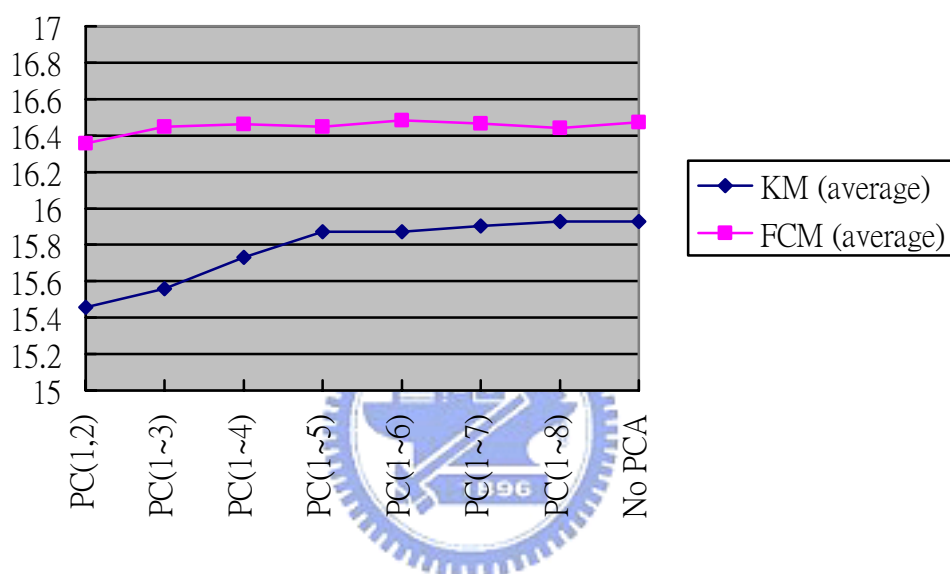


圖 5-11. 經 PCA 分析後取不同 feature 個數，K-Means 及 Fuzzy C-Means 分類結果的正確隊數曲線圖.

未經 PCA 分析的 2006 世界盃足球賽的統計資料分成 3 類的實驗結果列於表 5-14，從表中我們可以清楚地看出一個趨勢就是：屬於強隊的類別(A)及類別(B)，與屬於弱隊的類別(D)明顯地被分開到不同的 2 個類別中，而少數的類別(B)及大部份的類別(C)之球隊則交雜組合在一起成爲第 3 類。

取 4 個 features 經 PCA 分析的 2006 世界盃足球賽的統計資料分成 3 類的實驗結果列於表 5-15，從表中我們可以清楚地看出一個趨勢就是：屬於強隊的類別(A)及類別(B)，與屬於弱隊的類別(D)明顯地被分開到不同的 2 個類別中，而少數的類別(B)及大部份的類別(C)之球隊則交雜組合在一起成爲第 3 類。

未經 PCA 分析的 2006 世界盃足球賽的統計資料分成 2 類的實驗結果列於表 5-16，從表中我們可以清楚地看出一個趨勢就是：屬於弱隊的類別(D)明顯地被分開到 1 類中，而另外 1 類則由屬於強隊的類別(A)、類別(B)、及少數類別(C)之球隊組合而成。

取 4 個 features 經 PCA 分析的 2006 世界盃足球賽的統計資料分成 2 類的實驗結果列於表 5-17，從表中我們可以清楚地看出一個趨勢就是：屬於弱隊的類別(D)明顯地被分開到 1 類中，而另外 1 類則由屬於強隊的類別(A)、類別(B)、及少數類別(C)之球隊組合而成。



表 5-14. 未經 PCA 分析之 2006 世界盃足球賽的統計資料分成 3 類的分類結果.

Clustering Algorithm	Class	Clustering Result
KM	Class C1	法國(A)、義大利(A)
	Class C2	葡萄牙(A)、德國(A)、英國(B)、阿根廷(B)、巴西(B)、西班牙(C)、瑞士(C)、厄瓜多爾(C)、瑞典(C)、荷蘭(C)、巴拉圭(D)、墨西哥(C)、捷克(D)
	Class C3	烏克蘭(B)、加納(C)、哥斯大黎加(D)、象牙海岸(D)、波蘭(D)、千里達(D)、安哥拉(D)、伊朗(D)、美國(D)、澳大利亞(C)、日本(D)、突尼西亞(D)、沙特阿拉伯(D)、塞爾維亞(D)、多哥(D)、克羅地亞(D)、韓國(D)
FCM	Class C1	葡萄牙(A)、德國(A)、義大利(A)、英國(B)、阿根廷(B)、巴西(B)、西班牙(C)、瑞士(C)、厄瓜多爾(C)、瑞典(C)、荷蘭(C)、墨西哥(C)
	Class C2	法國(A)、烏克蘭(B)、加納(C)、澳大利亞(C)、韓國(D)、捷克(D)、巴拉圭(D)、象牙海岸(D)
	Class C3	哥斯大黎加(D)、波蘭(D)、千里達(D)、安哥拉(D)、美國(D)、日本(D)、突尼西亞(D)、塞爾維亞(D)、多哥(D)、沙特阿拉伯(D)、克羅地亞(D)、伊朗(D)
Hierarchical – single link	Class C1	法國(A)
	Class C2	象牙海岸(D)
	Class C3	德國(A)、義大利(A)、葡萄牙(A)、巴西(B)、英國(B)、烏克蘭(B)、阿根廷(B)、西班牙(C)、墨西哥(C)、瑞典(C)、荷蘭(C)、加納(C)、澳大利亞(C)、瑞士(C)、厄瓜多爾(C)、巴拉圭(D)、哥斯大黎加(D)、安哥拉(D)、沙特阿拉伯(D)、突尼西亞(D)、日本(D)、多哥(D)、捷克(D)、波蘭(D)、伊朗(D)、克羅地亞(D)、韓國(D)、美國(D)、千里達(D)、塞爾維亞(D)
Hierarchical – complete link	Class C1	德國(A)、義大利(A)、英國(B)、巴西(B)、西班牙(C)、瑞士(C)
	Class C2	法國(A)、象牙海岸(D)、韓國(D)、澳大利亞(C)、烏克蘭(B)、加納(C)、捷克(D)、墨西哥(C)、葡萄牙(A)、巴拉圭(D)、瑞典(C)、阿根廷(B)、荷蘭(C)、厄瓜多爾(C)
	Class C3	沙特阿拉伯(D)、塞爾維亞(D)、日本(D)、美國(D)、突尼西亞(D)、安哥拉(D)、千里達(D)、伊朗(D)、哥斯大黎加(D)、多哥(D)、克羅地亞(D)、波蘭(D)
Hierarchical – average link	Class C1	法國(A)、象牙海岸(D)
	Class C2	西班牙(C)、阿根廷(B)、荷蘭(C)、厄瓜多爾(C)、瑞士(C)、巴西(B)、義大利(A)、葡萄牙(A)、英國(B)、德國(A)
	Class C3	沙特阿拉伯(D)、多哥(D)、日本(D)、塞爾維亞(D)、韓國(D)、澳大利亞(C)、烏克蘭(B)、加納(C)、巴拉圭(D)、捷克(D)、墨西哥(C)、瑞典(C)、克羅地亞(D)、波蘭(D)、美國(D)、安哥拉(D)、突尼西亞(D)、千里達(D)、伊朗(D)、哥斯大黎加(D)
SOFM – 1D	Class C1	法國(A)、義大利(A)、德國(A)、葡萄牙(A)、英國(B)、巴西(B)、阿根廷(B)、西班牙(C)、荷蘭(C)、瑞士(C)
	Class C2	烏克蘭(B)、墨西哥(C)、厄瓜多爾(C)、瑞典(C)、加納(C)、澳大利亞(C)、象牙海岸(D)、巴拉圭(D)、捷克(D)、韓國(D)
	Class C3	美國(D)、安哥拉(D)、多哥(D)、波蘭(D)、克羅地亞(D)、千里達(D)、伊朗(D)、塞爾維亞(D)、哥斯大黎加(D)、突尼西亞(D)、日本(D)、沙特阿拉伯(D)

表 5-15. 取 4 個 features 經 PCA 分析後之 2006 世界盃足球賽的統計資料分成 3 類的分類結果.

Clustering Algorithm	Class	Clustering Result
KM	Class C1	德國(A)、葡萄牙(A)、義大利(A)、英國(B)、巴西(B)、瑞士(C)、西班牙(C)
	Class C2	法國(A)、阿根廷(B)、烏克蘭(B)、厄瓜多爾(C)、瑞典(C)、荷蘭(C)、加納(C)、墨西哥(C)、韓國(D)、巴拉圭(D)、象牙海岸(D)
	Class C3	澳大利亞(C)、波蘭(D)、哥斯大黎加(D)、千里達(D)、塞爾維亞(D)、安哥拉(D)、伊朗(D)、捷克(D)、美國(D)、克羅地亞(D)、日本(D)、多哥(D)、突尼西亞(D)、沙特阿拉伯(D)
FCM	Class C1	德國(A)、葡萄牙(A)、義大利(A)、英國(B)、阿根廷(B)、巴西(B)、西班牙(C)、瑞士(C)、厄瓜多爾(C)、墨西哥(C)、荷蘭(C)
	Class C2	法國(A)、瑞典(C)、巴拉圭(D)、加納(C)、捷克(D)、澳大利亞(C)、克羅地亞(D)、韓國(D)、烏克蘭(B)、沙特阿拉伯(D)、哥斯大黎加(D)
	Class C3	波蘭(D)、千里達(D)、象牙海岸(D)、塞爾維亞(D)、安哥拉(D)、伊朗(D)、美國(D)、日本(D)、多哥(D)、突尼西亞(D)
Hierarchical – single link	Class C1	法國(A)
	Class C2	象牙海岸(D)
	Class C3	德國(A)、義大利(A)、葡萄牙(A)、巴西(B)、英國(B)、烏克蘭(B)、阿根廷(B)、西班牙(C)、墨西哥(C)、瑞典(C)、荷蘭(C)、加納(C)、澳大利亞(C)、瑞士(C)、厄瓜多爾(C)、巴拉圭(D)、哥斯大黎加(D)、安哥拉(D)、沙特阿拉伯(D)、突尼西亞(D)、日本(D)、多哥(D)、捷克(D)、波蘭(D)、伊朗(D)、克羅地亞(D)、韓國(D)、美國(D)、千里達(D)、塞爾維亞(D)
Hierarchical – complete link	Class C1	美國(D)、安哥拉(D)、突尼西亞(D)、千里達(D)、哥斯大黎加(D)、伊朗(D)
	Class C2	法國(A)、象牙海岸(D)、瑞士(C)、巴西(B)、義大利(A)、西班牙(C)、英國(B)、德國(A)、葡萄牙(A)、阿根廷(B)、荷蘭(C)、厄瓜多爾(C)
	Class C3	烏克蘭(B)、墨西哥(C)、瑞典(C)、加納(C)、澳大利亞(C)、巴拉圭(D)、沙特阿拉伯(D)、日本(D)、多哥(D)、捷克(D)、波蘭(D)、克羅地亞(D)、韓國(D)、塞爾維亞(D)、
Hierarchical – average link	Class C1	法國(A)、象牙海岸(D)
	Class C2	美國(D)、千里達(D)、伊朗(D)、沙特阿拉伯(D)、塞爾維亞(D)、突尼西亞(D)、多哥(D)、安哥拉(D)、日本(D)、哥斯大黎加(D)、波蘭(D)、厄瓜多爾(C)、加納(C)、墨西哥(C)、澳大利亞(C)、瑞典(C)、荷蘭(C)、捷克(D)、克羅地亞(D)、韓國(D)、巴拉圭(D)、烏克蘭(B)
	Class C3	葡萄牙(A)、義大利(A)、德國(A)、英國(B)、巴西(B)、阿根廷(B)、西班牙(C)、瑞士(C)
SOFM – 1D	Class C1	義大利(A)、法國(A)、葡萄牙(A)、德國(A)、巴西(B)、英國(B)、阿根廷(B)、荷蘭(C)、墨西哥(C)、瑞士(C)、西班牙(C)、象牙海岸(D)
	Class C2	烏克蘭(B)、厄瓜多爾(C)、瑞典(C)、加納(C)、捷克(D)、澳大利亞(C)、韓國(D)、巴拉圭(D)
	Class C3	波蘭(D)、哥斯大黎加(D)、千里達(D)、塞爾維亞(D)、安哥拉(D)、伊朗(D)、美國(D)、克羅地亞(D)、日本(D)、多哥(D)、突尼西亞(D)、沙特阿拉伯(D)



表 5-16. 未經 PCA 分析之 2006 世界盃足球賽的統計資料分成 2 類的分類結果.

Clustering Algorithm	Class	Clustering Result
KM	Class C1	德國(A)、葡萄牙(A)、義大利(A)、法國(A)、英國(B)、巴西(B)、阿根廷(B)、烏克蘭(B)、厄瓜多爾(C)、瑞典(C)、荷蘭(C)、墨西哥(C)、瑞士(C)、澳大利亞(C)、西班牙(C)、象牙海岸(D)、捷克(D)
	Class C2	加納(C)、波蘭(D)、哥斯大黎加(D)、巴拉圭(D)、千里達(D)、塞爾維亞(D)、安哥拉(D)、伊朗(D)、美國(D)、克羅地亞(D)、日本(D)、多哥(D)、突尼西亞(D)、沙特阿拉伯(D)、韓國(D)
FCM	Class C1	義大利(A)、葡萄牙(A)、德國(A)、法國(A)、英國(B)、阿根廷(B)、烏克蘭(B)、巴西(B)、荷蘭(C)、瑞典(C)、墨西哥(C)、瑞士(C)、厄瓜多爾(C)、西班牙(C)、捷克(D)
	Class C2	加納(C)、澳大利亞(C)、波蘭(D)、哥斯大黎加(D)、巴拉圭(D)、千里達(D)、塞爾維亞(D)、安哥拉(D)、伊朗(D)、美國(D)、克羅地亞(D)、日本(D)、多哥(D)、突尼西亞(D)、沙特阿拉伯(D)、象牙海岸(D)、韓國(D)
Hierarchical – single link	Class C1	象牙海岸(D)
	Class C2	法國(A)、德國(A)、義大利(A)、葡萄牙(A)、巴西(B)、英國(B)、烏克蘭(B)、阿根廷(B)、西班牙(C)、墨西哥(C)、瑞典(C)、荷蘭(C)、加納(C)、澳大利亞(C)、瑞士(C)、厄瓜多爾(C)、巴拉圭(D)、哥斯大黎加(D)、安哥拉(D)、沙特阿拉伯(D)、突尼西亞(D)、日本(D)、多哥(D)、捷克(D)、波蘭(D)、伊朗(D)、克羅地亞(D)、韓國(D)、美國(D)、千里達(D)、塞爾維亞(D)
Hierarchical – complete link	Class C1	德國(A)、義大利(A)、葡萄牙(A)、法國(A)、英國(B)、巴西(B)、阿根廷(B)、烏克蘭(B)、西班牙(C)、瑞士(C)、澳大利亞(C)、墨西哥(C)、加納(C)、瑞典(C)、荷蘭(C)、厄瓜多爾(C)、象牙海岸(D)、韓國(D)、捷克(D)、巴拉圭(D)
	Class C2	沙特阿拉伯(D)、塞爾維亞(D)、日本(D)、美國(D)、突尼西亞(D)、安哥拉(D)、千里達(D)、伊朗(D)、哥斯大黎加(D)、多哥(D)、克羅地亞(D)、波蘭(D)
Hierarchical – average link	Class C1	德國(A)、法國(A)、義大利(A)、葡萄牙(A)、英國(B)、阿根廷(B)、巴西(B)、西班牙(C)、荷蘭(C)、厄瓜多爾(C)、瑞士(C)、象牙海岸(D)
	Class C2	烏克蘭(B)、加納(C)、墨西哥(C)、澳大利亞(C)、瑞典(C)、沙特阿拉伯(D)、多哥(D)、日本(D)、塞爾維亞(D)、韓國(D)、巴拉圭(D)、捷克(D)、克羅地亞(D)、波蘭(D)、美國(D)、安哥拉(D)、突尼西亞(D)、千里達(D)、伊朗(D)、哥斯大黎加(D)
SOFM – 1D	Class C1	義大利(A)、法國(A)、德國(A)、葡萄牙(A)、巴西(B)、阿根廷(B)、英國(B)、厄瓜多爾(C)、瑞典(C)、荷蘭(C)、墨西哥(C)、西班牙(C)、瑞士(C)、巴拉圭(D)
	Class C2	烏克蘭(B)、加納(C)、澳大利亞(C)、波蘭(D)、哥斯大黎加(D)、千里達(D)、象牙海岸(D)、塞爾維亞(D)、安哥拉(D)、伊朗(D)、捷克(D)、美國(D)、克羅地亞(D)、日本(D)、多哥(D)、突尼西亞(D)、沙特阿拉伯(D)、韓國(D)

表 5-17. 取 4 個 features 經 PCA 分析後之 2006 世界盃足球賽的統計資料分成 2 類的分類結果.

Clustering Algorithm	Class	Clustering Result
KM	Class C1	德國(A)、葡萄牙(A)、義大利(A)、法國(A)、英國(B)、巴西(B)、阿根廷(B)、烏克蘭(B)、瑞典(C)、荷蘭(C)、墨西哥(C)、加納(C)、澳大利亞(C)、瑞士(C)、西班牙(C)、象牙海岸(D)、捷克(D)、韓國(D)、厄瓜多爾(C)
	Class C2	波蘭(D)、哥斯大黎加(D)、巴拉圭(D)、千里達(D)、塞爾維亞(D)、安哥拉(D)、伊朗(D)、美國(D)、克羅地亞(D)、日本(D)、多哥(D)、突尼西亞(D)、沙特阿拉伯(D)
FCM	Class C1	義大利(A)、葡萄牙(A)、德國(A)、法國(A)、英國(B)、阿根廷(B)、烏克蘭(B)、巴西(B)、荷蘭(C)、瑞典(C)、墨西哥(C)、加納(C)、澳大利亞(C)、瑞士(C)、西班牙(C)、象牙海岸(D)、韓國(D)、捷克(D)
	Class C2	厄瓜多爾(C)、波蘭(D)、哥斯大黎加(D)、巴拉圭(D)、千里達(D)、塞爾維亞(D)、安哥拉(D)、伊朗(D)、美國(D)、克羅地亞(D)、日本(D)、多哥(D)、突尼西亞(D)、沙特阿拉伯(D)
Hierarchical – single link	Class C1	法國(A)
	Class C2	象牙海岸(D)、德國(A)、義大利(A)、葡萄牙(A)、巴西(B)、英國(B)、烏克蘭(B)、阿根廷(B)、西班牙(C)、墨西哥(C)、瑞典(C)、荷蘭(C)、加納(C)、澳大利亞(C)、瑞士(C)、厄瓜多爾(C)、巴拉圭(D)、哥斯大黎加(D)、安哥拉(D)、沙特阿拉伯(D)、突尼西亞(D)、日本(D)、多哥(D)、捷克(D)、波蘭(D)、伊朗(D)、克羅地亞(D)、韓國(D)、美國(D)、千里達(D)、塞爾維亞(D)
Hierarchical – complete link	Class C1	法國(A)、葡萄牙(A)、德國(A)、義大利(A)、阿根廷(B)、英國(B)、巴西(B)、西班牙(C)、瑞士(C)、象牙海岸(D)
	Class C2	烏克蘭(B)、厄瓜多爾(C)、瑞典(C)、荷蘭(C)、加納(C)、澳大利亞(C)、墨西哥(C)、巴拉圭(D)、克羅地亞(D)、捷克(D)、波蘭(D)、波蘭(D)、韓國(D)、沙特阿拉伯(D)、塞爾維亞(D)、日本(D)、美國(D)、突尼西亞(D)、安哥拉(D)、千里達(D)、伊朗(D)、哥斯大黎加(D)、多哥(D)
Hierarchical – average link	Class C1	法國(A)、葡萄牙(A)、義大利(A)、德國(A)、英國(B)、巴西(B)、阿根廷(B)、西班牙(C)、瑞士(C)、象牙海岸(D)
	Class C2	烏克蘭(B)、厄瓜多爾(C)、加納(C)、墨西哥(C)、澳大利亞(C)、瑞典(C)、荷蘭(C)、美國(D)、千里達(D)、伊朗(D)、沙特阿拉伯(D)、塞爾維亞(D)、突尼西亞(D)、多哥(D)、安哥拉(D)、日本(D)、哥斯大黎加(D)、波蘭(D)、捷克(D)、克羅地亞(D)、韓國(D)、巴拉圭(D)
SOFM – 1D	Class C1	德國(A)、葡萄牙(A)、義大利(A)、法國(A)、巴西(B)、英國(B)、阿根廷(B)、厄瓜多爾(C)、瑞典(C)、巴拉圭(C)、荷蘭(C)、墨西哥(C)、瑞士(C)、西班牙(C)、韓國(D)
	Class C2	烏克蘭(B)、加納(C)、波蘭(D)、哥斯大黎加(D)、千里達(D)、象牙海岸(D)、塞爾維亞(D)、安哥拉(D)、伊朗(D)、捷克(D)、美國(D)、澳大利亞(C)、克羅地亞(D)、日本(D)、多哥(D)、突尼西亞(D)、沙特阿拉伯(D)

## 6. 結論

本論文的主要貢獻為利用非監督式分類法中的 K-Means 分類演算法，Fuzzy C-Means 分類演算法，Hierarchical 分類演算法，及 SOFM 類神經網路，將 2006 年世界盃足球賽的球隊作分類，選取的特徵值有：(1)進球數 (Goals For (GF))，(2)射門次數 (Shots (S))，(3)射門命中球門次數 (Shots On Goal (SOG))，(4)角球數 (Corner Kicks (CK))，(5)直接自由球命中球門數 (Direct Free Kicks to Goal (DFKG))，(6)間接自由球命中球門數 (Indirect Free Kicks to Goal (IDFKG))，(7)持球時間 (Ball Possession (BP))，(8)被犯規次數 (Fouls Suffered (FS))共 8 項，根據競賽規則的特性我們把全部 32 隊參賽球隊分成 4 類，分類結果顯示各種非監督式分類法的分類正確率介於 46 % ~ 68.75 %之間。

進一步我們再利用以 Basic Sequential Clustering Algorithm、Modified Basic Sequential Clustering Algorithm 為基礎的有效性驗證法、及以 K-Means為基礎的 Pseudo F-Statistic 公式，共 3 種方法來作最佳分類個數的驗證，針對 2006 年世界盃足球賽統計數據進行驗證的結果顯示，2006 年世界盃足球賽官方統計數據的最佳分類個數以分成 2 類或 3 類為最佳之分類個數，同時我們把 2006 年世界盃足球賽官方統計數據先經過 PCA 分析後，再把原始資料投影到前四個主向量基底上，然後重新分成 2 類及 3 類的情況下，實驗結果可以明顯地看出實力強的球隊與實力弱的球隊被區分開。

## 參考文獻


- [1] 黃國源、張生平、張文龍、董安晉、及陳楷儒, “以類神經網路分析 2006 世界盃足球賽球隊實力”, 研究報告, 1 - 4 頁, 六月八日, 2006.
- [2] The match schedule, matches, results, and statistics reports of 2006 FIFA world cup Germany, <http://fifaworldcup.yahoo.com/06/en/>.
- [3] Official website of FIFA, <http://www.fifa.com/en/index.html>.
- [4] Sergios Theodoridis and Konstantinos Koutroumbas, Pattern Recognition, 3rd Edition, Academic Press, Inc., New York, 2006.
- [5] C. T. Lin and C. S. G. Lee, Neural Fuzzy Systems: A Neuro-Fuzzy Synergism to Intelligent Systems, Prentice Hall, 1996, 797 pages.
- [6] Kou-Yuan Huang, Neural Networks and Pattern Recognition, Weikeg Publishing Co., Taipei, Taiwan, March 2003, 406 pages.
- [7] Rui Xu and Donald Wunsch II, “Survey of clustering algorithm,” IEEE Transactions on Neural Networks, Vol.16, No.3, 2005, pp. 645-678.
- [8] E. W. Forgy, “Cluster analysis of multivariate data efficiency vs interpretability of classifications,” Biometrics 21, 1965, pp. 768-769.
- [9] J. C. Bezdek, Fuzzy Mathematics in Pattern Classification, PhD Thesis, Cornell University, Ithaca, NY. 1973.
- [10] T. Kohonen, H. Riittinen, E. Reuhkala, and S. Haltsonen, “On-line recognition of spoken words from a large vocabulary,” Information Sciences Journal, Vol. 33, Issue 1-2, 1984, pp. 3-30.
- [11] T. Kohonen, “Self-organized formation of topologically correct feature maps,” Neurocomputing, 1988, pp. 509-521.
- [12] T. Kohonen, “The elf-organizing map,” Proceedings of the IEEE, Vol.78, N0.9, 1990, pp. 1464-1480.
- [13] S. Kaski and T. Kohonen, “Winner-take-all networks for physiological models of competitive learning,” Neural Networks, Vol. 7, Issue 6-7, 1994, pp. 973-984.
- [14] T. Kohonen, “Physiological interpretation of the self-organizing map algorithm,” Neural Networks, Vol. 6, Issue 7, 1993, pp. 895-905.
- [15] A. V. Hall, "Methods for demonstrating resemblance in taxonomy and ecology," Nature, Vol. 214, 1967, pp. 830-831.
- [16] M. A. Vogel and A. K. C. Wong, “PFS clustering method,” IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol, PAMI-1, No. 3, 1979, pp. 237-245.

## 第二章 多層感知器類神經網路於 2006 年世界盃足球賽之勝率預測

### 摘要

本論文提出一個根據前一階段足球比賽之官方的統計數據，來預測下一階段兩隊比賽的勝率，這個足球勝率預測模型係利用以誤差倒傳遞學習演算法 (BP) 為基礎的多層感知器 (MLP) 的類神經網路，其 MLP 為 8-11-1 的網路架構，且 BP 的參數 momentum coefficient  $\beta$ 、learning rate  $\eta$ 、及 Mean Square Error (MSE) 經由實驗測試多組數據而決定後，用來預測 2006 年世界盃足球賽的勝率準確率達 62.5%，若排除比賽時間終了平手的比賽場次，則本文提出的方法準確率達到 76.9%。

### 1. 介紹



2006 年世界盃足球賽在德國舉行，從官方網站(<http://fifaworldcup.yahoo.com/06/en/>) [1] 中，我們可獲得由「世界足球聯盟組織 (FIFA) (<http://www.fifa.com/en/index.html>)」 [2] 針對球隊比賽之統計而得到各項足球技術的數據資料，例如：球隊進球數、失球數、控球時間、十二碼罰球數、紅牌數...等等，共計 17 項，每一個數據皆代表一項技術指標，而這些技術指標對球賽最終結果都具有影響力，因此針對這些技術指標，我們思考如何利用類神經網路，來預測球賽的勝負結果，1996 年 Michael C. Purucker 曾利用監督式及非監督式的類神經網路來分析預測 NFL 的勝率 [3]，J. Park and M. E. J. Newman 利用 one-parameter 的網路表示方法，針對 American College Football 的球隊作排名分析 [4]，還有 A. P. Rotshtein, M. Posner, 及 A. B. Rakityanskaya 結合基因演算及 Fuzzy Model 的類神經網路用來預測芬蘭的足球賽 [5]，根據 Michael C. Purucker 的論文研究結果 [3] 顯示監督式的訓練學習法比非監督式的訓練學習法可以得到更好的預測準確率，因此我們提出利用監督式多層感知器類神經網路及誤差倒傳遞 (Back Propagation) 學習演算法，來預測 2006 年世界盃足球賽兩隊比賽的勝負。我們採用的監督式圖形識別系統如圖 1 所示，方法為 MLP 及 BP learning rule，此系統可分訓練學習及分類辨識兩個部份，而我們要解決的問題就是利用前一階段的比賽統計資料來訓練這個網路之後，再用經過學習的 MLP 來預測下一階段兩隊比賽的球賽勝率。

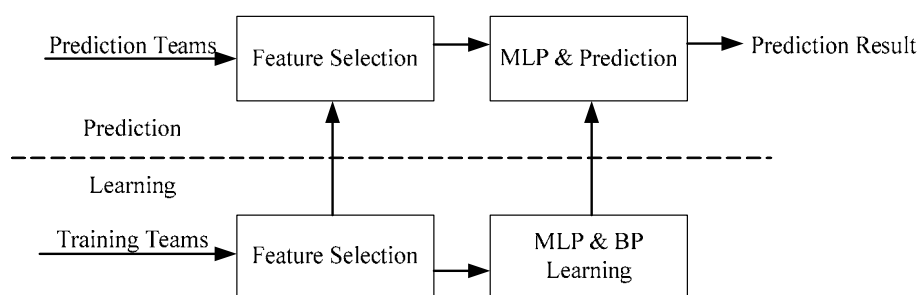


圖 1. Supervised Prediction System.

根據 2006 年世界盃足球賽賽程表，如圖 2 及附錄一 [1] 所示，本次進入會內賽共有 32 支隊伍，從比賽開始到比賽結束共有 64 場比賽，這 64 場比賽分成五個階段，每一階段的競賽規則說明如下：

第一階段 32 隊小組循環賽 (Group Matches)：此階段為不延長加時的比賽，在正規時間 90 分鐘比賽結束後記錄兩隊之比賽結果，總共有 32 隊及 48 場比賽，此階段為小組循環賽 (round robin tournament)，它把進入會內賽的 32 支隊伍分成 8 個小組 (Group A ~ Group H)，每小組內各有 4 隊，同小組內不同兩隊彼此會比賽一次作循環賽，所以每個小組各有 6 場比賽記錄，每支隊伍和自己有關的比賽記錄有 3 場，所以第一階段全部的比賽場次總數計有  $8 \times 6 = 48$  場 (Match 1~Match 48)，記分方法為勝一場得 3 分，平手則各得 1 分，敗一場則得 0 分，因此第一階段比賽結束後，取每小組中積分最高的前 2 名 (8 個小組中各取 2 名，共 16 隊) 進入第二階段比賽，第一階段比賽完後的績分如圖 3 所示，其中打 v 表示第一輪選取的訓練隊伍的積分為 9 (全勝) 或 0 (全敗)。

第二階段 16 隊單淘汰賽 (Round of 16)：此階段為單淘汰賽 (single elimination tournament)，在正規時間 90 分鐘終了若兩隊戰成平手，則進入延長加時賽 30 分鐘 (15 分鐘時場地互換)，若加時賽時間終了兩隊仍然平手，則進入十二碼罰球 (Penalty Kick) 大戰，直到分出勝負為止，本階段的隊伍由第一階段晉級的 16 隊角逐，共計 8 場比賽 (Match 49~Match 56)，勝者晉級下一階段，敗者則淘汰出局。

第三階段 8 隊四分之一決賽 (Quarter-finals)：此階段為單淘汰賽 (single elimination tournament)，勝負規則如同第二階段所述，本階段由第二階段晉級的 8 隊角逐，共計 4 場比賽 (Match 57~Match 60)，勝者晉級下一階段，敗者則淘汰出局。

第四階段 4 隊半決賽 (Semi-finals)：此階段為決賽的資格賽，勝負規則如同第二階



段所述，本階段由第三階段晉級的 4 隊角逐，共計 2 場比賽 (Match 61~Match 62)，勝者取得下一階段「冠亞軍賽 (final game)」的資格，敗者將取得下一階段「季軍賽 (third place game)」的資格。

第五階段 4 隊決賽 (Finals)：有 4 隊共 2 場比賽，一場為「季軍賽 (Match 63)」，另一場為「冠亞軍賽 (Match 64)」，勝負規則如同第二階段所述，其中季軍賽由第四階段的 2 支敗隊比賽，勝者為季軍 (第 3 名)，敗者為殿軍 (第 4 名)，冠亞軍賽由第四階段的 2 支勝隊比賽，勝者為冠軍 (第 1 名)，敗者為亞軍 (第 2 名)。

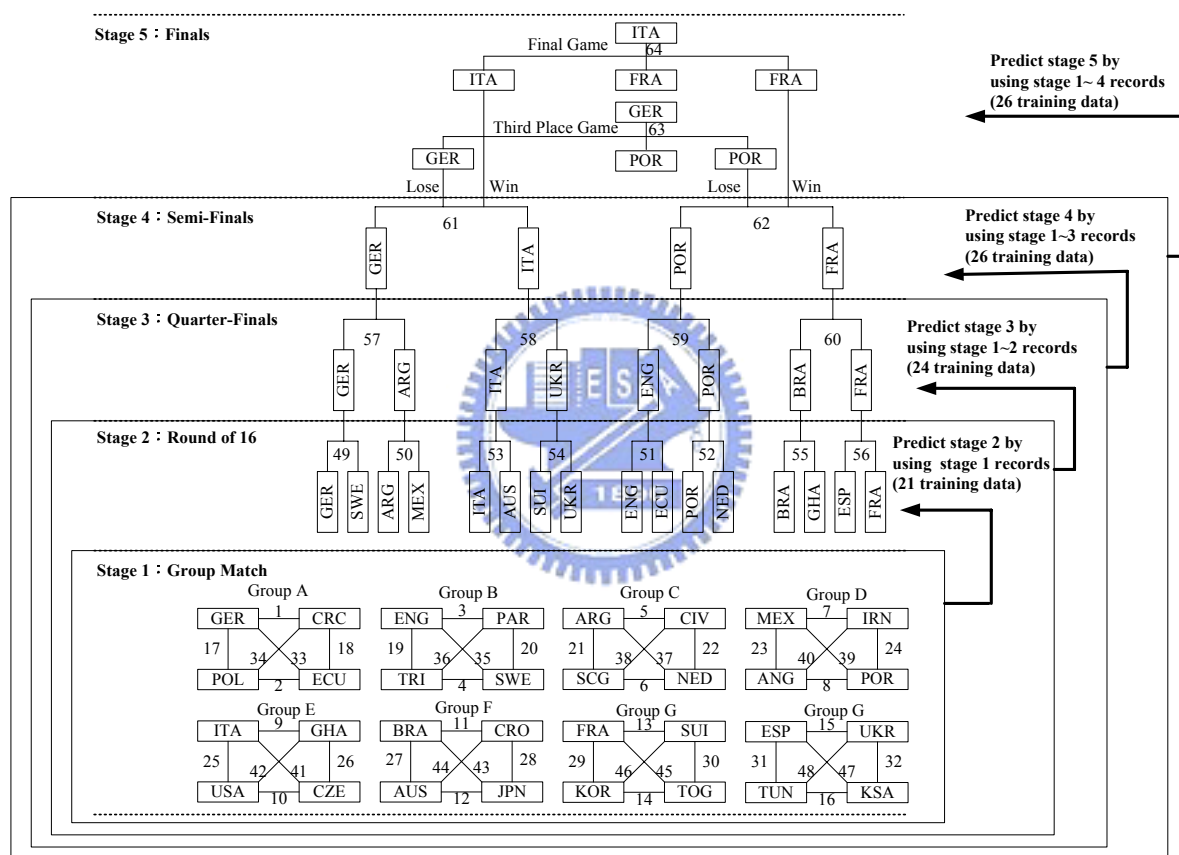


圖 2. 2006 年世界盃足球賽的訓練資料及預測的階段之關係圖。

Note: 32 隊參賽隊伍為 Germany (GER), Ecuador (ECU), Poland (POL), CRC (Costa Rica), ENG (England), SWE (Sweden), TRI (Trinidad and Tobago), PAR (Paraguay), ARG (Argentina), NED (Netherlands), CIV (Cote d'Ivoire), SCG (Serbia-Montenegro), MEX (Mexico), POR (Portugal), ANG (Angola), IRN (Iran), ITA (Italy), CZE (Czech Republic), USA (United States), GHA (Ghana), BRA (Brazil), JPN (Japan), AUS (Australia), CRO (Croatia), FRA (France), TOG (Togo), KOR (South Korea), SUI (Switzerland), ESP (Spain), KSA (Saudi Arabia), TUN (Tunisia), UKR (Ukraine) .



KEYS = W-Wins | L-Losses | D-Draws | GP-Games Played | GF-Goals For | GA-Goals against | WP-Winning % | Pt-Points

Group A							
Team	W	L	D	GP	GF	GA	Pt
✓ Germany	3	0	0	3	8	2	1.000 (9)
Ecuador	2	1	0	3	5	3	.667 6
Poland	1	2	0	3	2	4	.333 3
✓ Costa Rica	0	3	0	3	3	9	.000 (0)

Group B							
Team	W	L	D	GP	GF	GA	Pt
England	2	0	1	3	5	2	.667 7
Sweden	1	0	2	3	3	2	.333 5
Paraguay	1	2	0	3	2	2	.333 3
Trinidad and Tobago	0	2	1	3	0	4	.000 1

Group C							
Team	W	L	D	GP	GF	GA	Pt
Argentina	2	0	1	3	8	1	.667 7
Netherlands	2	0	1	3	3	1	.667 7
Cote d'Ivoire	1	2	0	3	5	6	.333 3
✓ Serbia-Montenegro	0	3	0	3	2	10	.000 (0)

Group D							
Team	W	L	D	GP	GF	GA	Pt
✓ Portugal	3	0	0	3	5	1	1.000 (9)
Mexico	1	1	1	3	4	3	.333 4
Angola	0	1	2	3	1	2	.000 2
Iran	0	2	1	3	2	6	.000 1

Group E							
Team	W	L	D	GP	GF	GA	Pt
Italy	2	0	1	3	5	1	.667 7
Ghana	2	1	0	3	4	3	.667 6
Czech Republic	1	2	0	3	3	4	.333 3
United States	0	2	1	3	2	6	.000 1

Group F							
Team	W	L	D	GP	GF	GA	Pt
✓ Brazil	3	0	0	3	7	1	1.000 (9)
Australia	1	1	1	3	5	5	.333 4
Croatia	0	1	2	3	2	3	.000 2
Japan	0	2	1	3	2	7	.000 1

Group G							
Team	W	L	D	GP	GF	GA	Pt
Switzerland	2	0	1	3	4	0	.667 7
France	1	0	2	3	3	1	.333 5
South Korea	1	1	1	3	3	4	.333 4
✓ Togo	0	3	0	3	1	6	.000 (0)

Group H							
Team	W	L	D	GP	GF	GA	Pt
✓ Spain	3	0	0	3	8	1	1.000 (9)
Ukraine	2	1	0	3	5	4	.667 6
Tunisia	0	2	1	3	3	6	.000 1
Saudi Arabia	0	2	1	3	2	7	.000 1

KEYS = W-Wins | L-Losses | D-Draws | GP-Games Played | GF-Goals For | GA-Goals against | WP-Winning % | Pt-Points



圖 3. 2006 年世界盃足球賽第一階段積分圖。

## 2. 多層感知器(MLP)及倒傳遞(BP)學習演算法

### 2.1 多層感知器(MLP)網路架構

David Rumelhart、Geoffrey Hinton、及 Ronald Williams [6] 三人於 1986 年提出多層感知器(Multi-Layer Perceptron)，簡稱 MLP，以及著名的倒傳遞(Back Propagation)學習演算法，簡稱 BP，如圖 4 所示為一個 8-11-1 的 MLP，而 BP 演算法在 Rumelhart 及 McClelland 的 PDP (Parallel Distributed Processing) [7] 一書中收錄而得到普及。

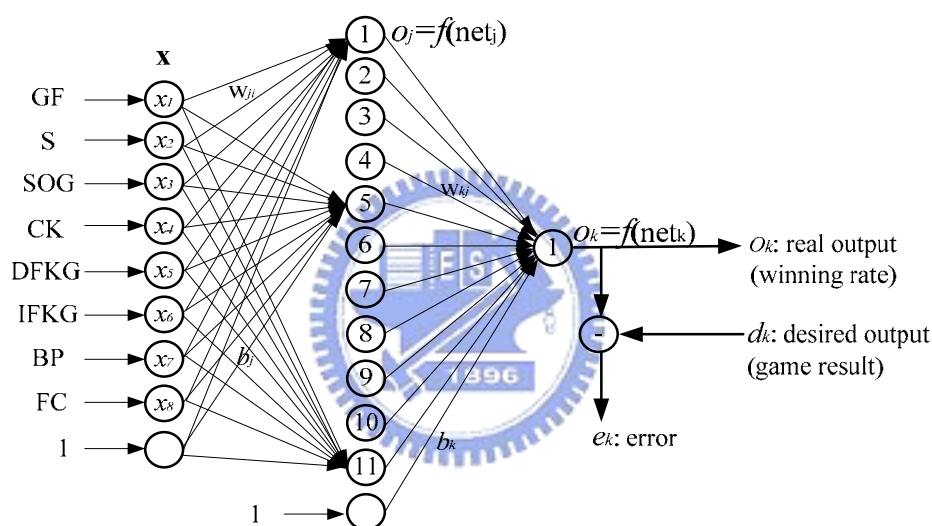


圖 4. 預測 2006 年世界盃足球賽勝率所採用的 MLP 網路架構。

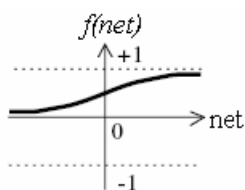
### 2.2 倒傳遞(BP)學習規則

如圖 4 的 8-11-1 三層 (一層隱藏層) MLP， $\mathbf{x}$  為輸入向量， $w_{ji}$  及  $w_{kj}$  為連結權重， $b$  為偏權值， $\text{net}$  為淨輸出， $f(\text{net})$  為轉換函數， $o$  為輸出值， $d$  為期望值， $e$  為誤差值。

我們採用的 8-11-1 MLP 網路架構中的隱藏層及輸出層的轉換函數皆採用 Logsigmoid function，公式為：

$$f(\text{net}) = \frac{1}{1 + e^{-\text{net}}}$$

Logsigmoid function 輸出及輸入的關係圖為：



而微分  $f'(net) = f(net)(1 - f(net))$  .

各參數符號之間的關係為

$$o_j = f(net_j) = f\left(\sum_{i=1}^8 w_{ji}x_i + b_j\right)$$

$$o_k = f(net_k) = f\left(\sum_{j=1}^{11} w_{kj}o_j + b_k\right)$$

倒傳遞學習訓練的過程及目的是在降低輸出值  $o$  與期望值  $d$  之間的誤差值  $e$ ，也就是最小化誤差函數  $E$  的過程，因此若誤差函數為

$$E = \sum_{k=1}^K e_k = \frac{1}{2} \sum_{k=1}^K (d_k - o_k)^2$$

我們利用 Gradient Decent 來求得  $E$  的最佳解，因此

$$w(t+1) = w(t) + \Delta w$$

$$\Delta w = -\eta \frac{\partial E}{\partial w}$$

其中  $\eta$  為學習速率，而  $\frac{\partial E}{\partial w}$  可用連鎖律得到，詳細的推導過程及通用公式請參閱 [8]。

以下以一個三層（一層隱藏層）MLP 為例子，若輸入層有  $I$  個神經元，隱藏層有  $J$  個神經元，輸出層有  $K$  個神經元，則各層之間權重係數的調整計算公式如下：

(1) 輸出層及隱藏層之間的權重係數調整公式：

$$\Delta w_{kj} = \eta(d_k - o_k)f'_k(net_k)o_j, \quad k=1, \dots, K, \quad \text{and} \quad j=1, \dots, J+1,$$

(2) 隱藏層及輸入層的權重係數調整公式：

$$\Delta w_{ji} = \eta \left[ \sum_{k=1}^K (d_k - o_k)f'_k(net_k)w_{kj} \right] f'_j(net_j)o_i, \quad j=1, \dots, J+1, \quad \text{and} \quad i=1, \dots, I$$

BP 演算過程中  $\eta$  的選擇對網路學習有很大的影響，若  $\eta$  設得太小，則學習速度將變得很慢，若  $\eta$  設得較大，雖可加快網路學習的速度，但由於使用 Gradient Decent 方法求最佳化時，不一定保證能得到 global minimum 的解，在學習過程中有可能會落入 local minimum，造成震盪，因此 Rumelhart, Hinton, 及 Williams [6] 建議加入動量項 (momentum coefficient  $\beta$ ) 來改善收斂過程中的震盪現象，加入  $\beta$  的主要作用是讓前一次權重調整  $\Delta w(t-1)$  的變化，對目前權重調整  $\Delta w(t)$  移動方向產生類似前一次調整的慣性 (inertia) 效果，同時保有一定程度的動量 (momentum)。

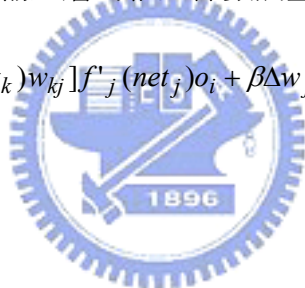
加入  $\beta$  後權重的調整量變為  $\Delta w^m(t) = \Delta w(t) + \beta \Delta w(t-1)$  [8] - [10]，因此修正後的權重係數調整計算公式如下：

(1) 加入慣性項  $\beta$  後，輸出層及隱藏層之間的權重係數調整公式：

$$\Delta w_{kj}(t) = \eta(d_k - o_k)f'_k(net_k)o_j + \beta\Delta w_{kj}(t-1)$$

(2) 加入慣性項  $\beta$  後，隱藏層及輸入層的權重係數調整公式：

$$\Delta w_{ji}(t) = \eta \left[ \sum_{k=1}^K (d_k - o_k)f'_k(net_k)w_{kj} \right] f'_j(net_j)o_i + \beta\Delta w_{ji}(t-1)$$



### 3. 2006 年世界盃足球賽的勝率預測

#### 3.1 原始資料的擷取與轉換

根據 2006 年世界盃足球賽的官方網頁 (<http://fifaworldcup.yahoo.com>) 我們可得到每一場比賽的比賽報告，如附錄二所示為德國(GER)對哥斯大黎加(CRC)之比賽報告，共計有 64 場，每場比賽報告的統計項目共計有 17 項，如表 1 所列，從這 17 個統計項目中，我們擷取其中可以代表球隊實力或影響球隊實力的 8 項統計數據做為輸入資料的特徵值，分別標示為  $x_1 \sim x_8$ ， $x_1$  為進球數 (Goals For (GF))， $x_2$  為射門次數 (Shots (S))， $x_3$  為射門命中球門次數 (Shots On Goal (SOG))， $x_4$  為角球數 (Corner Kicks (CK))， $x_5$  為直接自由球命中球門數 (Direct Free Kicks to Goal (DFKG))， $x_6$  為間接自由球命中球門數 (Indirect Free Kicks to Goal (IDFKG))， $x_7$  為持球時間 (Ball Possession (BP))， $x_8$  為被犯規次數 (Fouls Suffered (FS))， $x_1 \sim x_8$  在輸入 MLP 預測網路之前須先經過正規化處理，其中  $x_1 \sim x_8$  的值愈大代表球隊實力愈強，其輸入值的正規化計算為：

*If*  $x_{iA} = x_{iB}$ , *then*

$$y_{iA} = y_{iB} = 0.5$$

*Else*

$$y_{iA} = \frac{x_{iA}}{x_{iA} + x_{iB}}, \quad y_{iB} = \frac{x_{iB}}{x_{iA} + x_{iB}}, \quad i = 1, \dots, 8$$

$A$  代表 A 隊， $B$  代表 B 隊， $i$  為 8 個特徵值的索引，上述正規化計算後， $x_{iA}$  與  $x_{iB}$  的值將介於 0~1 之間。我們設 *If*  $x_{iA} = x_{iB}$  , *then*  $y_{iA} = y_{iB} = 0.5$ ，是因為如果 A 隊及 B 隊的進球數均為 0，則輸入值變為  $y_{iA} = 0.5$ ,  $y_{iB} = 0.5$ 。

德國(GER)對哥斯大黎加(CRC) 8 項原始資料的擷取及正規化轉換後的資料列於表 2。

表 1. 德國隊(GER)對哥斯大黎加隊(CRC) 17 項原始資料.

(技術指標)統計項目	簡寫	選為特徵值	A 球隊(GER)	B 球隊(CRC)
進球數 (Goals For)	GF	$x_1$	4	2
失球數 (Goal Against)	GA		2	4
射門次數 (Shots)	S	$x_2$	21	4
射門命中球門數 (Shots On Goal)	SOG	$x_3$	10	2
十二碼罰球數 (Penalty Kicks)	PK		0	0
犯規次數 (Fouls Committed)	FC		11	15
黃牌數 (Yellow Cards)	Y		0	1
紅牌數 (Red Cards)	R		0	0
角球數 (Corner Kicks)	CK	$x_4$	7	3
直接自由球命中球門數 (Direct Free Kicks to Goal)	DFKG	$x_5$	1	0
間接自由球命中球門數 (Indirect Free Kicks to Goal)	IDFKG	$x_6$	0	0
越位次數 (Offside)	O		3	3
烏龍球 (Own Goals)	OG		0	0
警告 (Cautions)	C		0	1
驅逐出場 (Expulsions)	E		0	0
持球時間 (Ball Possession)	BP	$x_7$	63%	37%
被犯規次數 (Fouls Suffered)	FS	$x_8$	12	11

Note: 1. 十二碼球次數只含正規時間及加時賽的部份，PK 大戰不列入統計.

2. GER 的「被犯規次數」與 CRC 的「犯規次數」不同，是因為犯規次數的記錄含非對人的犯規，例如：手球，邊線球.

表 2. 德國隊(GER)對哥斯大黎加隊(CRC) 8 項原始資料的擷取及正規化轉換.

(技術指標)統計項目	簡寫	特徵值	A 球隊(GER) 轉換前 $x_{iA}$	B 球隊(CRC) 轉換前 $x_{iB}$	A 球隊(GER) 轉換後 $y_{iA}$	B 球隊(CRC) 轉換後 $y_{iB}$
Goals For	GF	$x_1$	4	2	0.6666	0.3333
Shots	S	$x_2$	21	4	0.84	0.16
Shots On Goal	SOG	$x_3$	10	2	0.8333	0.1666
Corner Kicks	CK	$x_4$	7	3	0.7	0.3
Direct Free Kicks to Goal	DFKG	$x_5$	1	0	1	0
Indirect Free Kicks to Goal	IDFKG	$x_6$	0	0	0.5	0.5
Ball Possession	BP	$x_7$	63%	37%	0.63	0.37
Fouls Suffered	FS	$x_8$	12	11	0.5217	0.4782

### 3.2 網路訓練的目標值

因為我們選取第一階段的 training teams 為全勝或全敗，因此 desired output 為 1 或 0，第二階段之後加入的 training teams 均為勝隊，其 desired output 為 1。

### 3.3 訓練資料的選取及預測隊伍之輸入資料的轉換

如 2006 年世界盃足球賽的賽制所示比賽共分五個階段，我們的目的是用前面所有階段的統計資料來預測下一階段兩隊比賽的勝率，所以要預測的對象為後四階段(第二階段到第五階段)所有比賽場次的隊伍，因此訓練資料的選取及預測隊伍之輸入資料的轉換說明如下所述。

#### 訓練資料的選取

訓練資料選取的原則為：在第一階段我們只選取全勝或全敗的隊伍作為訓練資料，在第二階段的勝隊且其在第一階段為全勝者將其資料加入作為 training data，之後階段均依此原則來選取訓練資料，預測第二階段到第五階段的訓練資料選取結果如下所述。

預測第二階段 (Round of 16) 的訓練資料：我們選取第一階段 32 隊中，積分為 9 分 (3 場全勝) 的隊伍做為勝隊的訓練隊伍，0 分 (3 場全敗) 的隊伍做為敗隊的訓練隊伍，將這些勝隊與敗隊相關比賽場次的數據拿來做為訓練資料，因此可以找到 7 隊 (全勝隊為 GER, POR, BRA, ESP，全敗隊為 CRO, SCG, TOG) 共 21 組的訓練資料，訓練 MLP 後，預測第二階段共 8 場的比賽。

預測第三階段 (Quarter Finals) 的訓練資料：除第一階段的 21 組資料外，再加上有晉級第三階段 8 隊中，若其在第一階段為 3 場全勝者，則把其在第二階段的比賽記錄加進來，所以可以找到 3 隊 (GER, BRA, POR)，因此第三階段總計有  $21+3=24$  組的訓練資料，訓練 MLP 後，預測第三階段共 4 場的比賽。

預測第四階段 (Semi Finals) 的訓練資料：除之前階段的 24 組資料外，再加上有晉級第四階段 4 隊中，其在第一階段為全勝者，則把其在第三階段比賽的記錄加進來，所以可以找到 2 隊 (GER, POR)，所以第四階段總計有  $24+2=26$  組的訓練資料，訓練 MLP 後，預測第四階段共 2 場的比賽。

預測第五階段 (Finals) 的訓練資料：因為晉級第五階段冠亞軍戰的 2 隊 (FRA, ITA)，其在第一階段並非為全勝晉級的隊伍，所以第五階段的訓練資料與第四階段的訓練資料相同，共有 26 組訓練資料，訓練 MLP 後，預測第五階段最後 2 場的比賽。



## 預測隊伍之輸入資料

第一階段 (Group Matches)：第一階段不需預測，所以不需要預測的輸入資料。

預測第二階段 (Round of 16) 8 場比賽共 16 隊的輸入資料：我們取有晉級到第二階段的 16 隊中，把每隊在第一階段與其相關的前 3 場比賽全部加起來後，取 3 場平均值來做為該隊伍在第二階段的預測輸入資料，因此有 16 隊輸入資料作預測。如表 3 所列，為德國隊(A 隊)對瑞典隊(B 隊)在第二階段的預測輸入資料。

預測第三階段 (Quarter Finals) 4 場比賽共 8 隊的輸入資料：我們取有晉級到第三階段的 8 隊中，與這 8 隊本身相關的第 4 場 (第二階段) 比賽資料與其第一階段的 3 場比賽資料全部加起來後，取這 4 場的平均值來做為各隊的預測輸入資料。

預測第四階段 (Semi Finals) 2 場比賽共 4 隊的輸入資料：我們取有晉級到第四階段的 4 隊中，與這 4 隊本身相關的第 5 場 (第三階段) 比賽資料，與其在第一階段到第二階段的 4 場比賽資料全部加起來後，取這 5 場的平均值來做為各隊的預測輸入資料。

預測第五階段 (Finals) 2 場比賽共 4 隊的輸入資料：第五階段決賽的 4 隊和第四階段半決賽的 4 隊相同，差別在於每隊都會多 1 場比賽記錄，我們取與這 4 隊有關的第 6 場 (第四階段) 比賽資料與其在第一階段到第三階段的 5 場比賽資料全部加起來後，取這 6 場的平均值來做為這 4 隊的預測輸入資料。

2006 年世界盃足球賽的比賽結果及晉級的勝負關係，請參考附錄三的「比賽結果記錄表」，全部 64 場的比賽報告，請參考官方網站 (<http://fifaworldcup.yahoo.com/06/en/>)。

表 3. 預測德國隊(GER)對瑞典隊(SWE)的輸入資料.

統計項目	德國隊(GER)				瑞典隊(SWE)			
	第一場	第二場	第三場	三場平均	第一場	第二場	第三場	三場平均
$x_1$ (GF)	0.6664	1	1	0.8889	0.5	1	0.5	0.6667
$x_2$ (S)	0.84	0.7619	0.6818	0.7929	0.75	0.7692	0.4286	0.6493
$x_3$ (SOG)	0.8333	0.7619	0.8182	0.7612	0.75	0.5152	0.3913	0.5522
$x_4$ (CK)	0.7	0.7143	0.45	0.6214	0.4737	0.6667	0.6667	0.6023
$x_5$ (DFKG)	1	0.5	0	0.5	1	0.5	0.5	0.6667
$x_6$ (IDFKG)	0.5	0.5	0	0.3333	0.5	0.5	0.5	0.5
$x_7$ (BP)	0.63	0.58	0.43	0.5467	0.6	0.57	0.37	0.5400
$x_8$ (FS)	0.5217	0.5526	0.538	0.5374	0	0.4412	0.4194	0.2868

### 3.4 MLP 的 BP 學習法則參數的決定

本實驗利用 MATLAB 7.0 版中的 Neural Network Toolbox 之 newff 函數來建立一個 feed-forward backpropagation network，訓練及學習的演算法為“Gradient Descent with Momentum Back-Propagation”，為了決定此 MLP 之 hidden neurons 個數、及 BP 訓練學習法的 momentum coefficient  $\beta$  及 learning rate  $\eta$  三種參數的使用值，我們先利用預測第二階段的 21 筆訓練資料來訓練此 MLP，如表 4 所列為決定三種參數之使用值時，各種測試條件的固定值、測試條件的變數值、及觀察記錄的項目。

我們可藉由各種參數的設定及測試之結果圖中的觀察，來決定各項參數最後的選定值，詳細的實驗結果如下各小節所述。

表 4. 決定三種參數使用值之測試條件.

欲決定的參數	測試條件的固定值	測試條件的變數值	觀察記錄的項目
Nnumber of hidden neurons	(1) $\beta=0.5$ (2) $\eta=0.1$	(1) MSE = 0.1, 0.05, 0.01 (2) Hidden neurons 2~40	Hidden neurons 與 epochs 的關係
Momentum coefficient $\beta$	(1) Hidden neurons = 11 (2) MSE=0.01 (3) $\eta=0.1$	(1) $\beta = 0.5, 0.6, 0.7$ (2) Tesing times = 40	$\beta$ 與 epochs 的關係
Learning rate $\eta$	(1) Hidden neurons = 11 (2) MSE=0.01 (3) $\beta=0.6$	(1) $\eta = 0.1, 0.3, 0.5, 0.7, 0.9$ (2) Tesing times = 40	$\eta$ 與 epochs 的關係

## 隱藏層神經元個數的決定

依 V. Kecman 的實驗建議  $\beta=0.5 \sim 0.7$ ,  $0 < \eta < 1$  [11], 因此我們先將  $\beta$  及  $\eta$  選定在  $\beta=0.5$ ,  $\eta=0.1$ , 測試隱藏層神經元個數從 2 ~ 40 在 3 種不同訓練誤差 MSE (0.1, 0.05, 0.01) 下與 epochs 的關係, 由圖 5 之(a)~(c)的結果中我們發現不論神經元的個數從 2~40 個及訓練誤差為 0.1、0.05、及 0.01 之下, 網路的訓練學習皆能收斂, 但趨勢顯示較小的 MSE 需花費的收斂次數相對較多。另外我們將(a)~(c) 3 種不同 neurons 設定時, MSE 及 epochs 的關係劃在(d)圖中, 從(d)圖的觀察可看出在 3 種不同的 MSE 下, 當 neurons 數由 2 往上增加到超過 11 個時, epochs 的變化明顯地由快速下降變為緩慢, 因此隱藏層神經元的個數我們決定採用 11 個神經元。

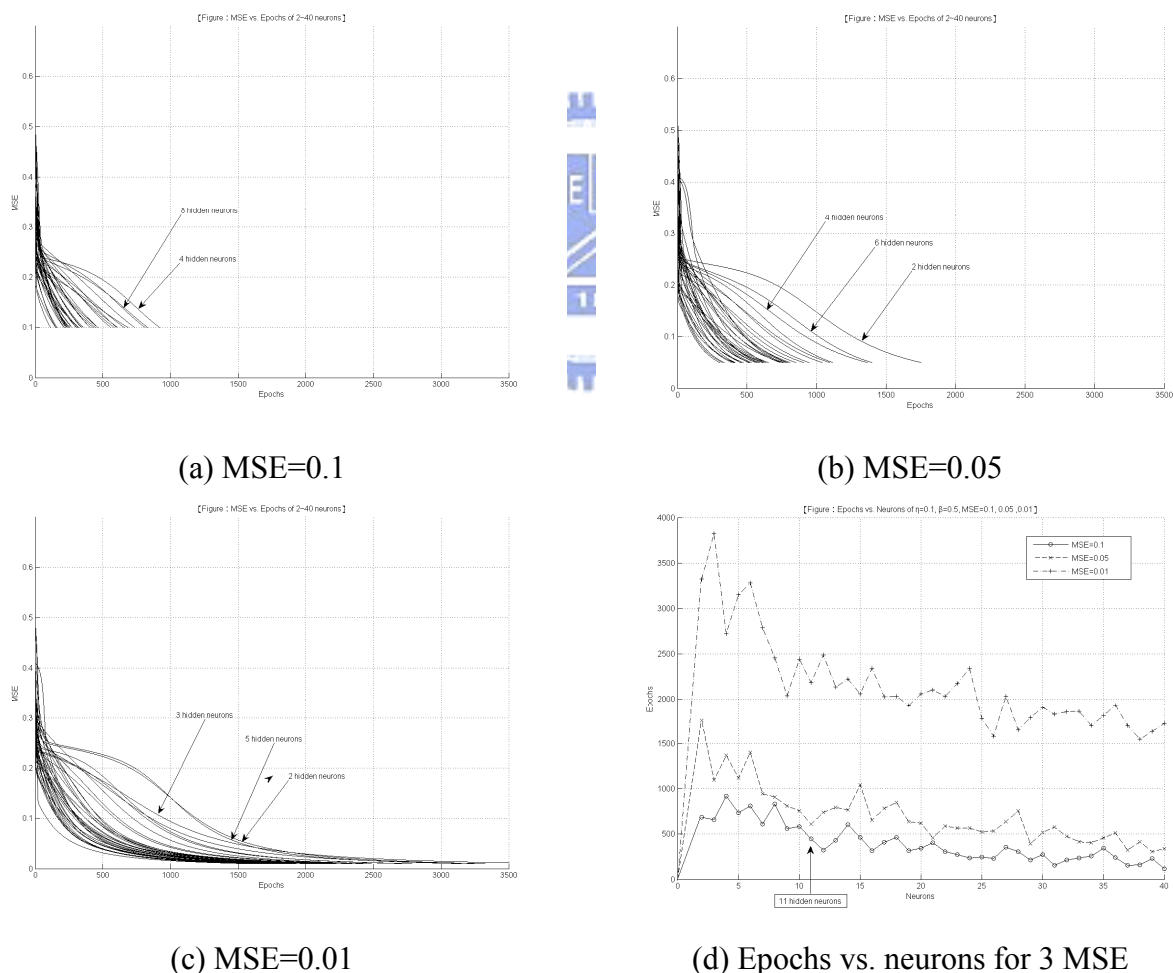


圖 5. 2 ~ 40 Hidden neurons 與 3 種 MSE 的 epochs 比較圖。

## Momentum coefficient $\beta$ 的決定

由上實驗結果我們設定 hidden neurons=11, MSE=0.01, 並選定較慢的學習速率 learning rate  $\eta=0.1$ , 然後測試 3 種不同  $\beta$  值 (0.5, 0.6, 0.7) 設定與 epochs 的關係, 每一個  $\beta$  設定各測試 40 次, 實驗結果顯示於圖 6 的(a)(b)(c), 此外我們把(a)(b)(c) 3 種  $\beta$  設定與 epochs 的對應關係結果劃在(d)圖中, 從(d)圖的觀察, 我們可以看出當  $\beta=0.6$  時, 針對不同隱藏層神經元個數而言, 其收斂的次數相對比較穩定, 因此決定 momentum coefficient  $\beta=0.6$ 。

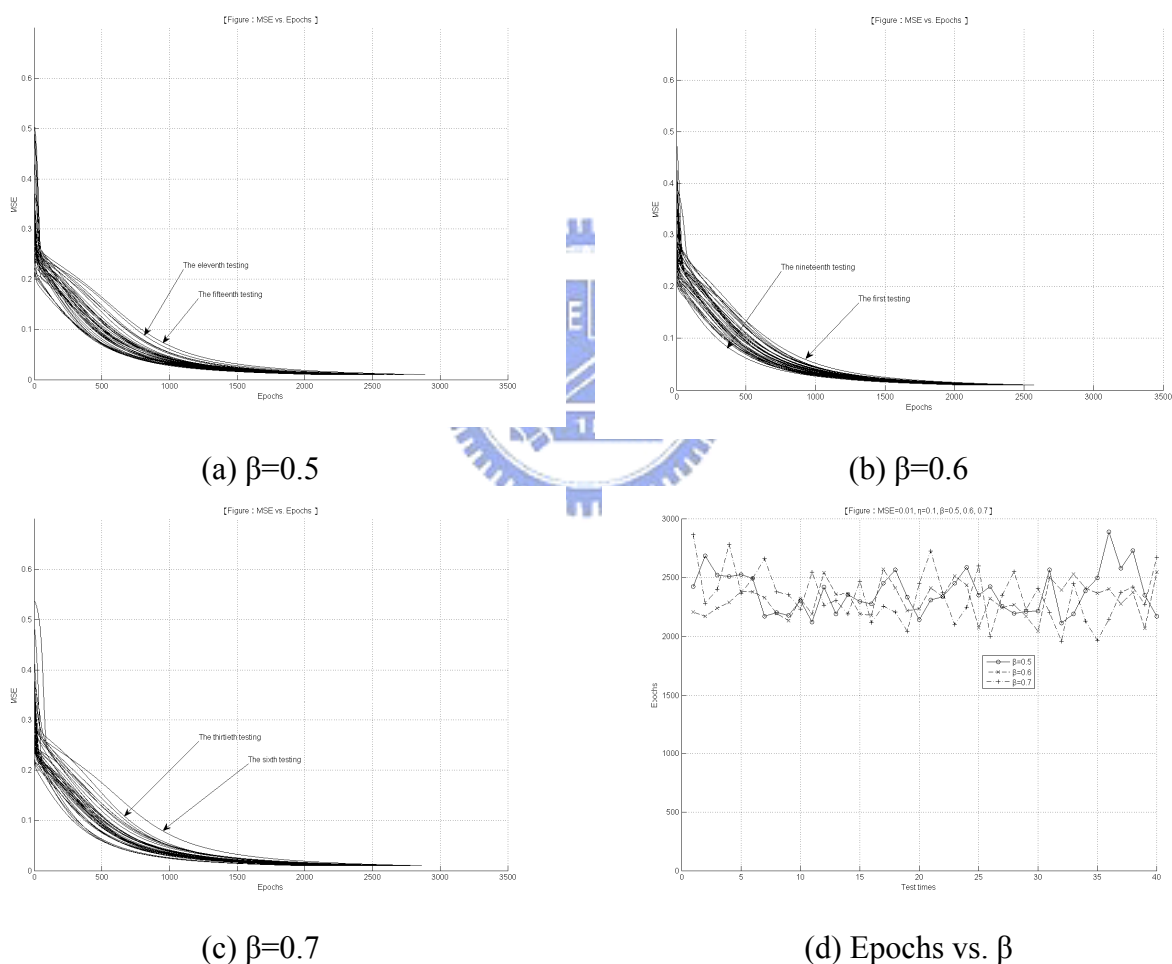


圖 6. 設定 MSE=0.01,  $\eta=0.1$ , hidden neurons=11, 測試 3 種  $\beta$  與 epochs 的比較圖。(每一個  $\beta$  設定各測試 40 次).

## Learning Rate $\eta$ 的決定

由以上實驗結果，我們選定 hidden neurons=11, MSE=0.01, 及 momentum coefficient  $\beta=0.6$ ，分別測試 5 種 learning rate  $\eta$  ( $\eta=0.1, 0.3, 0.5, 0.7, 0.9$ ) 設定與 epochs 的關係，每一個  $\eta$  設定各測試 40 次，實驗結果顯示於圖 7 的(a)~(e)，此外我們把(a)~(e) 5 種不同  $\eta$  設定與 epochs 的對應關係結果劃在(f)圖，從(f)圖的觀察，我們可以清楚地看出當  $\eta=0.9$  時，對網路學習而言都是最快收斂，因此選定 learning rate  $\eta=0.9$ 。

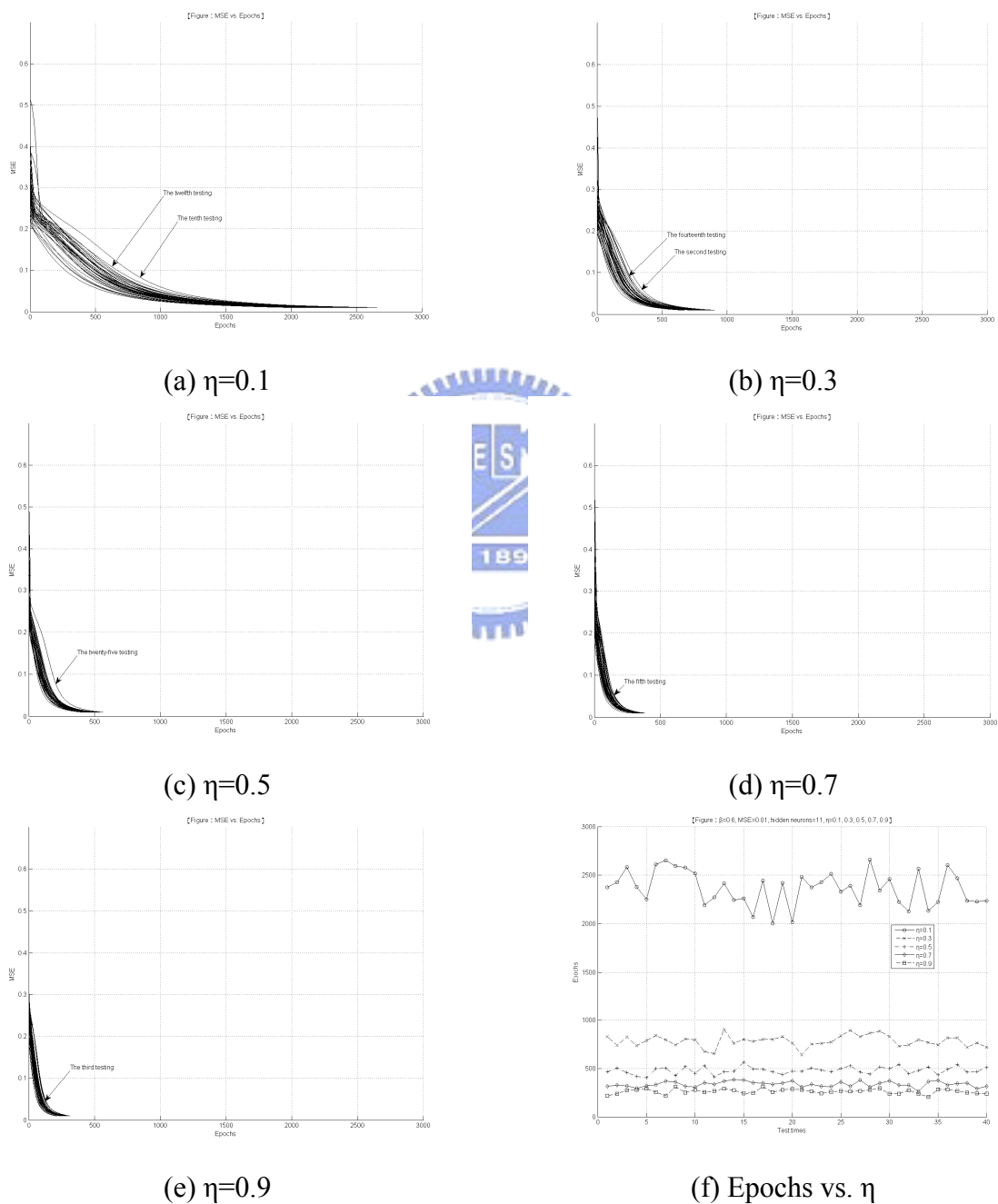
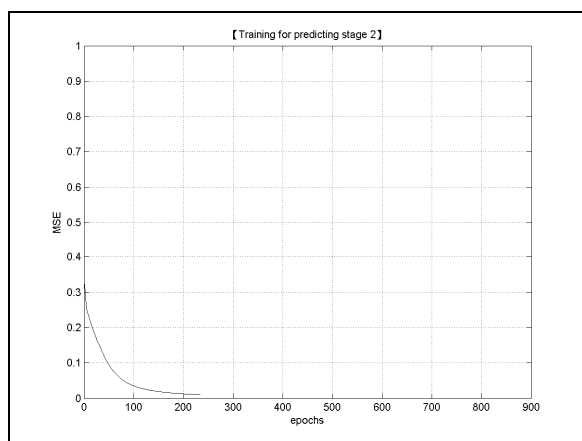


圖 7. 設定 MSE=0.01,  $\beta=0.6$ , hidden neurons=11, 測試 5 種  $\eta$  與 epochs 的比較圖. (每一個  $\eta$  設定各測試 40 次).

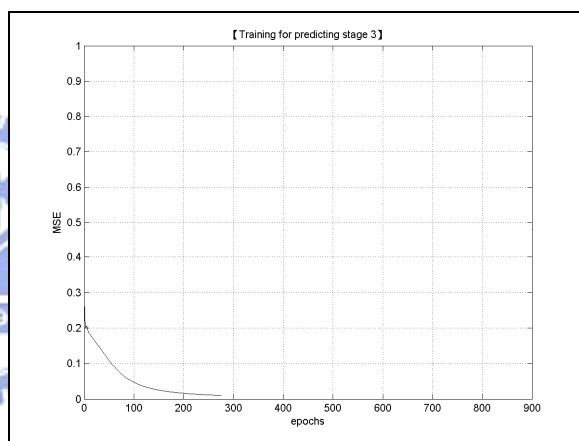
## 4. 2006 年世界盃足球賽勝率預測的結果與討論

### 4.1 勝率預測結果

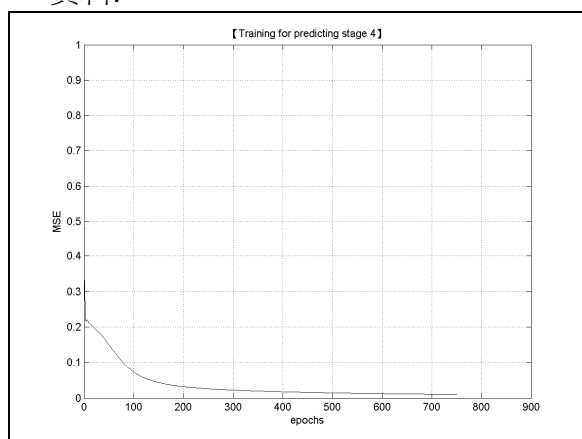
最後本實驗採用的網路模型為：8-11-1 MLP，學習演算法為 Gradient Decent with Momentum Back-Propagation，訓練過程之參數設定為： $\eta=0.9$ ,  $\beta=0.6$ ,  $MSE=0.01$ 。每階段的預測方法為將欲預測隊伍的平均輸入資料經由輸入層輸入到訓練過的網路，然後在其輸出層的輸出值即為該隊的勝率，我們經由比較兩隊的勝率的大小即可得到勝敗關係。預測第二階段到第五階段勝率的訓練過程如圖 8 所示，橫軸代表 epochs，縱軸代表 MSE。



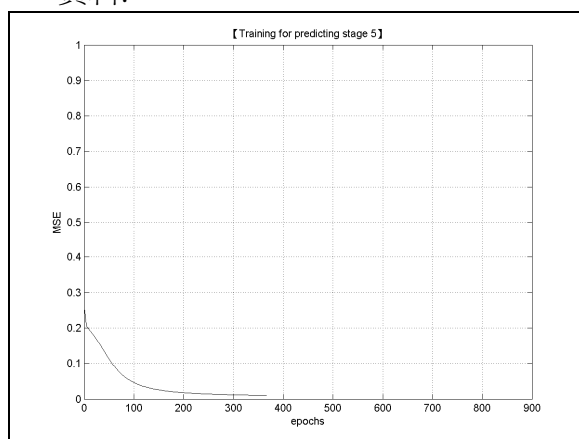
(a) 預測第二階段的訓練過程，共 21 筆訓練資料。



(b) 預測第三階段的訓練過程，共 24 筆訓練資料。



(c) 預測第四階段的訓練過程，共 26 筆訓練資料。



(d) 預測第五階段的訓練過程，共 26 筆訓練資料。

圖 8. MLP 的 BP 的訓練過程圖。

預測第二階段到第五階段勝率的實驗結果列於表 5，表中列出的項目有：比賽階段、比賽場次、比賽隊伍、網路輸出值、預測結果、比賽結果、預測是否正確、預測正確率、及經計算後的參考賠率供參考。其中：

比賽階段：代表第二階段到第五階段的比賽階段。

比賽場次：代表官方網站賽程表上的比賽場次。

網路輸出值：為利用網路預測某球隊的勝率的輸出值。

預測結果：網路輸出值較高的為“W”，網路輸出值較低的為“L”。

比賽結果：以官方提供的正規時間終了或加時賽時間終了的結果為結果，其為正規時間及加時賽時間終了結束後兩隊仍然平手就視同此兩隊為平手，而不以 PK 大戰之後的勝負結果為結果。勝負關係符號為：

“W”代表勝，“D”代表平手，“L”代表敗。

預測是否正確：預測該場次比賽的勝負關係是否正確，正確為“Y”，不正確為“N”。

預測正確率：該階段比賽的預測正確率。

參考賠率：賠率的計算公式為：

$$\text{Odds (Team A vs. Team B)} = \frac{y_B}{y_A}, \text{ 其中 } y_A \text{ 及 } y_B \text{ 為兩隊的網路輸出值}$$

例如：

若 Odds (Team A vs. Team B) = 1:1.5，代表 A 隊若獲勝則壓 1 賠 1.5。

從表 4 的實驗結果中我們可以看出，本論文提出的倒傳遞多層感知器類神經網路用於預測 2006 年世界盃足球賽的結果，在第二階段到第五階段的預測正確率各為：

預測第二階段 8 場的正確率：75% (6/8)

預測第三階段 4 場的正確率：50% (2/4)

預測第四階段 2 場的正確率：50% (1/2)

預測第五階段 2 場的正確率：50% (1/2)

其中值得注意的是在預測第二階段勝出隊伍時的預測正確率可達 75%，這樣的結果跟一些所謂「專家預測」的結果來比是令人滿意且可以接受的，更何況足球比賽勝負的預測在所有運動項目中是最難預測的一項，至於第三階段、第四階段、及第五階段的預測正確率就僅有 50%。若是不分階段把所有預測的比賽場次全部加起來只看整體預測正確率，那本研究提出的預測方法之正確率達到 62.5% (10/16)，這樣的結果也是可接受的。



表 5. 第二階段到第五階段的預測結果表.

比賽階段	比賽場次	隊伍	網路輸出	預測結果	比賽結果	預測是否正確	預測正確率(%)	參考賠率
第二階段	49	GER	0.93682	W	W	Y	75% (6/8)	0.9035
		SWE	0.84641	L	L			1.1068
	50	ARG	0.9004	W	W	Y		0.68883
		MEX	0.62023	L	L			1.4517
	51	ENG	0.93173	W	W	Y		0.6879
		ECU	0.64094	L	L			1.4537
	52	POR	0.74147	L	W	N		1.0785
		NED	0.79969	W	L			0.92721
	53	ITA	0.83955	W	W	Y		0.89024
		AUS	0.74741	L	L			1.1233
	54	SUI	0.93074	W	D	N		0.84312
		UKR	0.78472	L	D			1.1861
	55	BRA	0.93571	W	W	Y		0.74993
		GHA	0.70172	L	L			1.3335
	56	GER	0.95556	W	W	Y		0.9711
		SWE	0.92794	L	L			1.0298
第三階段	57	GER	0.95291	W	D	N	50% (2/4)	0.91223
		ARG	0.86927	L	D			1.0962
	58	ITA	0.93416	W	W	Y		0.78832
		UKR	0.73642	L	L			1.2685
	59	ENG	0.86278	L	L	Y		1.0667
		POR	0.92033	W	W			0.93746
60	BRA	0.9344	W	L	N	0.97236		
	GER	0.90857	L	W		1.0284		
第四階段	61	GER	0.98742	L	L	Y	50% (1/2)	1.0071
		ITA	0.99093	W	W			0.993
	62	POR	0.98098	W	L	N		0.9944
		FRA	0.98272	L	W			1.0056
第五階段	63	GER	0.85393	W	W	Y	50% (1/2)	0.96184
		POR	0.82134	L	L			1.0397
	64	ITA	0.95056	W	D	N		0.97134
		FRA	0.92332	L	D			1.0295

## 4.2 預測結果的討論

針對預測錯誤的場次，我們逐一檢視並分析如下：

第二階段預測錯誤的比賽：

第 52 場 葡萄牙 vs 荷蘭

這一場比賽兩隊都各有二名球員得到紅牌而出場，變成一場 9 對 9 的比賽，從比賽結果的數據看出整場的各项優勢都顯示荷蘭優於葡萄牙(與預測期望相同)，但葡萄牙卻藉由對手一個犯規獲得的直接自由球得分而獲勝，這是令人無法預測到的結果。

第 54 場 蘇格蘭 vs 烏克蘭

經由比賽結果中看出，這場比賽結果顯示蘇格蘭優於烏克蘭(與預測期望相同)，但此役最終的結果是平手。

第三階段預測錯誤的比賽：

第 57 場 德國 vs 阿根廷

經由比賽結果中看出，這場比賽結果顯示阿根廷略優於德國(與預測期望相同)，且阿根廷得到一張紅牌以 10 人對 11 人，若沒有這張紅牌阿根廷勝出的機會將大增，而這場比賽的結果是雙方平手。

第 60 場 巴西 vs 法國

從賽後數據看來，巴西隊應該優於法國隊，然而比賽結果由法國隊獲得勝利(與預測期望不同)，而巴西落敗也是事實，這場比賽和大多數人的預測結果確實相反，我們只能說“球是圓的，巴西隊失常的演出致使法國隊晉級”。

第四階段預測錯誤的比賽：

第 62 場 葡萄牙 vs 法國

從賽後數據看來，其實葡萄牙佔有相當多優勢，不論是控球時間還是射門次數皆優於法國隊(與預測期望相同)，無耐機會無法有效把握，卻在一次禁區內的犯規讓法國隊在正規時間內以一個十二碼罰球將葡萄牙淘汰出局。

第 64 場 義大利 vs 法國

從比賽結果中我們看出這場球賽雙方球隊各有優勢，互有領先，但義大利略優於法國(與預測期望相同)，然而這場比賽在正規時間結束時雙方戰成平手。

從上面的分析中，很明顯我們提出的 MLP 預測網路對於平手的場次無法有效地分辨出，究其原因如下：

- (1) 由於第一階段為 8 個小組的循環賽，然而從中晉級的隊伍在第一階段不是 3 戰全勝的隊伍，就是 2 勝 1 和或是 2 勝 1 負的隊伍，然而這些晉級的隊伍也是我們將在第二階段之後所要預測的隊伍，因此網路對這些隊伍預測的輸出值都偏向於 1 (大於 0.5)，然而由於本研究提出的方法為根據網路預測之輸出值的大小來判定勝負關係，所以對於平手的場次就無法有效地分辨。
- (2) 有鑑於此，改進的建議做法是，若是比賽的賽制改成所有參賽球隊全部是循環賽(例如歐洲足球俱樂部的聯賽制度)，彼此至少會比賽一次，那在這種循環賽賽制下，當強隊與弱隊碰上時，本研究提出的網路勝率預測輸出值將會得到更好的效果(更明確的分離)。

基於我們提出的 MLP 預測網路對於平手的場次無法有效地分辨出來，若是我們將比賽結果為平手的 3 個場次 (Match 54, 57, 64) 排除不列入正確率的計算，那麼整體預測的正確率將大為提高，如表 6 所列，可以看出整體預測之正確率將提高到 76.9 % (10/13)。

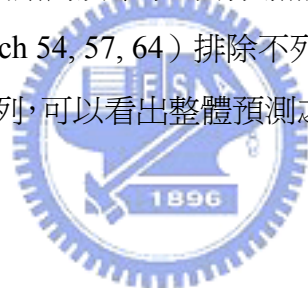


表 6. 第二階段到第五階段不含平手場次的預測正確率.

比賽階段	比賽場次	隊伍	網路輸出	預測結果	比賽結果	預測是否正確	預測正確率(%)	整體預測正確率
第二階段	49	GER	0.93682	W	W	Y	85.7 % (6/7)	76.9 % (10/13)
		SWE	0.84641	L	L			
	50	ARG	0.9004	W	W	Y		
		MEX	0.62023	L	L			
	51	ENG	0.93173	W	W	Y		
		ECU	0.64094	L	L			
	52	POR	0.74147	L	W	N		
		NED	0.79969	W	L			
	53	ITA	0.83955	W	W	Y		
		AUS	0.74741	L	L			
	54	SUI	0.93074	<del>W</del>	D	<del>X</del>		
		UKR	0.78472	<del>L</del>	D			
55	BRA	0.93571	W	W	Y			
	GHA	0.70172	L	L				
56	GER	0.95556	W	W	Y			
	SWE	0.92794	L	L				
第三階段	57	GER	0.95291	<del>W</del>	D	<del>X</del>	66.67 % (2/3)	
		ARG	0.86927	<del>L</del>	D			
	58	ITA	0.93416	W	W	Y		
		UKR	0.73642	L	L			
	59	ENG	0.86278	L	L	Y		
		POR	0.92033	W	W			
60	BRA	0.9344	W	L	N			
	GER	0.90857	L	W				
第四階段	61	GER	0.98742	L	L	Y	50 % (1/2)	
		ITA	0.99093	W	W			
	62	POR	0.98098	W	L	N		
FRA		0.98272	L	W				
第五階段	63	GER	0.85393	W	W	Y	100% (1/1)	
		POR	0.82134	L	L			
	64	ITA	0.95056	<del>W</del>	D	<del>X</del>		
		FRA	0.92332	<del>L</del>	D			

## 5. 結論

本研究主要建立一個 MLP 類神經網路，並用倒傳遞學習規則來訓練此 MLP 後，用來預測 2006 年世界盃足球賽兩隊比賽的勝率，若排除比賽平手的場次不計，我們的預測準確率可達 76.9 %，其中主要的關鍵在於我們從 2006 年世界盃足球賽的統計資料中，選取可以反應球隊實力的技術指標作為特徵值，其為：(1)進球數 (Goals For (GF))、(2)射門次數 (Shots (S))、(3)射門命中球門次數 (Shots On Goal (SOG))、(4)角球數 (Corner Kicks (CK))、(5)直接自由球命中球門數 (Direct Free Kicks to Goal (DFKG))、(6)間接自由球命中球門數 (Indirect Free Kicks to Goal (IDFKG))、(7)持球時間 (Ball Possession (BP))、及(8)被犯規次數 (Fouls Suffered (FS))共 8 項，經過正規化轉換後輸入 MLP，網路的大小為 8-11-1。基於世界盃的賽制，下一階段預測的訓練與預測過程由前面所有階段的比賽資料作訓練，再預測每一隊的預測勝率。在 BP learning 方面，我們做實驗得到一層 hidden layer 的最佳 hidden node 個數為 11 個，learning rate  $\eta = 0.9$ ，momentum coefficient  $\beta = 0.6$ 。

由於一場球賽欲做完整的各項技能統計需花費龐大的人力及成本，因此取得更充份完整的球賽統計數據來作為訓練資料及預測資料，將進一步提升預測的可靠性。同時一場球賽勝負的關鍵除了個人技術水平，整體戰力之基本條件外，有一些因素是我們無法掌控的，例如：運氣的成份，球員收賄打假球，天候因素，主客場心理因素，...，等等，但無論如何，本研究已利用現有的統計數據做最客觀的足球勝率預測，我們期望未來能激勵起更多運動科學的專家或運動愛好者，投注更多的心力在此方面的研究。

## 參考文獻

- [1] 2006 FIFA world cup Germany – match schedule, matches and results, and statistics reports, <http://fifaworldcup.yahoo.com/06/en/>.
- [2] Official website of FIFA, <http://www.fifa.com/en/index.html>.
- [3] C. P. Michael, “Neural networks quarterbacking-how different training methods perform in calling the games,” IEEE Potentials, August/September 1996, pp.9-15.
- [4] J. Park and M. E. J. Newman, “A network-based ranking system for US college football,” Journal of Statistical Mechanics: Theory and Experiment, IOP publishing, Issue 10, 2005, pp.1-14.
- [5] A. P. Rotshtein, M. Posner, and A. B. Rakityyanskaya, “Football predictions based on a fuzzy model with genetic and neural tuning,” Cybernetics and Systems Analysis, Vol.41, No.4, 2005, pp.619-630.
- [6] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, “Learning representations by backpropagating errors,” Nature, Vol.233, 1986, pp.533-536.
- [7] David E. Rumelhart and James L. McClelland, Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol.1, MIT Press: Cambridge, MA, 1986.
- [8] Kou-Yuan Huang, Neural Networks and Pattern Recognition, Weikey Publishing Co., Taipei, Taiwan, March 2003, 406 pages.
- [9] R. A. Jacobs, “Increase rates of convergence through learning rate adaptation,” Neural Networks, Vol.1, 1988, pp. 295-307.
- [10] J. M. Zurada, “Lambda learning rule for feedforward neural networks,” Proceedings of the IEEE International Conference in Neural Networks, Vol.3, 1993, pp.1808-1811.
- [11] V. Kecman, Learning and Soft Computing: Support Vector Machines, Neural Networks, and Fuzzy Logic Models, MIT Press: Cambridge, MA, 2001.



附錄一. 2006 世界盃足球賽賽程表.



2006 FIFA World Cup Germany™ - Match Schedule

First Round	09.06.06 Friday	10.06.06 Saturday	11.06.06 Sunday	12.06.06 Monday	13.06.06 Tuesday	14.06.06 Wednesday	15.06.06 Thursday	16.06.06 Friday	17.06.06 Saturday	18.06.06 Sunday	19.06.06 Monday	20.06.06 Tuesday	21.06.06 Wednesday	22.06.06 Thursday	23.06.06 Friday
Berlin					Match 11 BRA - CRD 21:30 hrs		Match 20 SWE - PAR 21:30 hrs						Match 33 ECU - GER 18:30 hrs		Match 48 USA - TUR 18:30 hrs
Dortmund		Match 4 TRI - GWE 18:30 hrs				Match 17 GEO - POL 21:00 hrs					Match 20 TGO - GAI 15:30 hrs			Match 43 JPN - BRA 21:00 hrs	
Frankfurt			Match 3 ENG - PAR 18:30 hrs		Match 14 KOR - TGO 15:30 hrs				Match 26 KOR - KOR 15:30 hrs					Match 37 MEX - ARG 21:30 hrs	
Gelsenkirchen	Match 2 POL - ECU 21:30 hrs			Match 15 USA - GBR 18:30 hrs				Match 27 ARG - SCO 15:30 hrs					Match 38 FRA - MEX 18:30 hrs		
Hamburg		Match 5 ARG - CIV 21:30 hrs					Match 18 ECU - CRC 18:30 hrs				Match 32 KSA - UCR 18:30 hrs			Match 41 CZE - ITA 18:30 hrs	
Hanover				Match 8 ITA - GHA 21:30 hrs				Match 23 MEX - ARG 21:30 hrs				Match 34 CRC - POL 18:30 hrs			Match 46 SLO - KOR 21:00 hrs
Kaiserslautern				Match 12 AUS - JPN 15:30 hrs					Match 25 ITA - USA 21:30 hrs				Match 36 PAR - TRK 21:30 hrs		Match 47 KSA - ESP 18:30 hrs
Cologne			Match 6 ARG - POR 21:30 hrs						Match 28 CZE - GHA 18:30 hrs				Match 39 SWE - SLO 21:30 hrs		Match 45 TGO - FRA 21:00 hrs
Leipzig			Match 9 SCO - NED 18:30 hrs			Match 15 ESP - UCR 15:30 hrs							Match 40 ESP - ARG 18:30 hrs		
Munich	Match 1 GER - CRC 18:30 hrs					Match 16 TUR - KSA 18:30 hrs				Match 27 BRA - AUS 18:30 hrs			Match 39 CIV - SCO 21:30 hrs		Match 42 GHA - USA 18:00 hrs
Nuremberg			Match 7 MEX - DEN 18:30 hrs				Match 18 SLO - TRK 18:30 hrs			Match 28 JPN - CRC 18:30 hrs				Match 42 GHA - USA 18:00 hrs	
Stuttgart					Match 13 FRA - GAI 18:30 hrs			Match 22 NED - CIV 18:30 hrs			Match 31 ESP - TUR 21:30 hrs			Match 44 CRC - AUS 21:30 hrs	

Second Round	Round of 16				Quarter-finals		Semi-finals		3rd Place	Final
Venue	24.06.06 Saturday	25.06.06 Sunday	26.06.06 Monday	27.06.06 Tuesday	30.06.06 Friday	01.07.06 Saturday	04.07.06 Tuesday	05.07.06 Wednesday	08.07.06 Saturday	09.07.06 Sunday
Berlin					Match 57 VEN - VEN 17:30 hrs					Match 64 SWE - MEX 20:00 hrs
Dortmund				Match 58 IR - SVK 17:00 hrs			Match 61 RSC - RSC 21:00 hrs			
Frankfurt						Match 60 RSC - RSC 21:30 hrs				
Gelsenkirchen						Match 59 VEN - VEN 17:30 hrs				
Hamburg					Match 59 VEN - VEN 17:30 hrs					
Hanover				Match 56 UR - SVK 21:00 hrs						
Kaiserslautern			Match 53 UR - SVK 17:00 hrs							
Cologne			Match 54 UR - SVK 21:00 hrs							
Leipzig	Match 53 UR - SVK 21:00 hrs									
Munich	Match 48 UR - SVK 17:30 hrs						Match 62 VEN - VEN 21:00 hrs			
Nuremberg		Match 52 UR - SVK 21:00 hrs								
Stuttgart		Match 51 UR - SVK 17:00 hrs							Match 63 UR - SVK 21:00 hrs	

Endgültige Fassung · Final version · Versión definitiva · Versión definitiva

Sat 10 Dec 2005; 01:10

Official Partners of the 2006 FIFA World Cup Germany™

Official Suppliers of the 2006 FIFA World Cup Germany™

(from <http://fifaworldcup.yahoo.com>)



## 2006 FIFA World Cup Germany™ - Match Schedule

<b>Group A</b>
Germany
Costa Rica
Poland
Ecuador
<b>Group B</b>
England
Paraguay
Trinidad and Tobago
Sweden
<b>Group C</b>
Argentina
Côte d'Ivoire
Serbia & Montenegro
Netherlands
<b>Group D</b>
Mexico
Iran
Angola
Portugal
<b>Group E</b>
Italy
Ghana
USA
Czech Republic
<b>Group F</b>
Brazil
Croatia
Australia
Japan
<b>Group G</b>
France
Switzerland
Korea Republic
Togo
<b>Group H</b>
Spain
Ukraine
Tunisia
Saudi Arabia

### Group Matches

Date	Match	Venue	Time	Teams
Fri 09.06.06	1	Munich	18:00 hrs	Germany - Costa Rica
	2	Geisenkirchen	21:00 hrs	Poland - Ecuador
Sat 10.06.06	3	Frankfurt	15:00 hrs	England - Paraguay
	4	Dortmund	18:00 hrs	Trinidad and Tobago - Sweden
	5	Hamburg	21:00 hrs	Argentina - Côte d'Ivoire
Sun 11.06.06	6	Leipzig	15:00 hrs	Serbia & Montenegro - Netherlands
	7	Nuremberg	18:00 hrs	Mexico - Iran
	8	Cologne	21:00 hrs	Angola - Portugal
Mon 12.06.06	12	Kaiserslautern	15:00 hrs	Australia - Japan
	10	Geisenkirchen	18:00 hrs	USA - Czech Republic
	9	Hanover	21:00 hrs	Italy - Ghana
Tue 13.06.06	14	Frankfurt	15:00 hrs	Korea Republic - Togo
	13	Stuttgart	18:00 hrs	France - Switzerland
	11	Berlin	21:00 hrs	Brazil - Croatia
Wed 14.06.06	15	Leipzig	15:00 hrs	Spain - Ukraine
	16	Munich	18:00 hrs	Tunisia - Saudi Arabia
	17	Dortmund	21:00 hrs	Germany - Poland
Thu 15.06.06	18	Hamburg	15:00 hrs	Ecuador - Costa Rica
	19	Nuremberg	18:00 hrs	England - Trinidad and Tobago
	20	Berlin	21:00 hrs	Sweden - Paraguay
Fri 16.06.06	21	Geisenkirchen	15:00 hrs	Argentina - Serbia & Montenegro
	22	Stuttgart	18:00 hrs	Netherlands - Côte d'Ivoire
	23	Hanover	21:00 hrs	Mexico - Angola
Sat 17.06.06	24	Frankfurt	15:00 hrs	Portugal - Iran
	26	Cologne	18:00 hrs	Czech Republic - Ghana
	25	Kaiserslautern	21:00 hrs	Italy - USA
Sun 18.06.06	28	Nuremberg	15:00 hrs	Japan - Croatia
	27	Munich	18:00 hrs	Brazil - Australia
	29	Leipzig	21:00 hrs	France - Korea Republic
Mon 19.06.06	30	Dortmund	15:00 hrs	Togo - Switzerland
	32	Hamburg	18:00 hrs	Saudi Arabia - Ukraine
	31	Stuttgart	21:00 hrs	Spain - Tunisia
Tue 20.06.06	33	Berlin	16:00 hrs	Ecuador - Germany
	34	Hanover	16:00 hrs	Costa Rica - Poland
	36	Kaiserslautern	21:00 hrs	Paraguay - Trinidad and Tobago
	35	Cologne	21:00 hrs	Sweden - England
Wed 21.06.06	40	Leipzig	16:00 hrs	Iran - Angola
	39	Geisenkirchen	16:00 hrs	Portugal - Mexico
	37	Frankfurt	21:00 hrs	Netherlands - Argentina
	38	Munich	21:00 hrs	Côte d'Ivoire - Serbia & Montenegro
Thu 22.06.06	41	Hamburg	16:00 hrs	Czech Republic - Italy
	42	Nuremberg	16:00 hrs	Ghana - USA
	43	Dortmund	21:00 hrs	Japan - Brazil
	44	Stuttgart	21:00 hrs	Croatia - Australia
Fri 23.06.06	48	Berlin	16:00 hrs	Ukraine - Tunisia
	47	Kaiserslautern	16:00 hrs	Saudi Arabia - Spain
	45	Cologne	21:00 hrs	Togo - France
	46	Hanover	21:00 hrs	Switzerland - Korea Republic

### Round of 16

Date	Match	Venue	Time	Teams
Sat 24.06.06	49	Munich	17:00 hrs	1A - 2B
	50	Leipzig	21:00 hrs	1C - 2D
Sun 25.06.06	51	Stuttgart	17:00 hrs	1B - 2A
	52	Nuremberg	21:00 hrs	1D - 2C
Mon 26.06.06	53	Kaiserslautern	17:00 hrs	1E - 2F
	54	Cologne	21:00 hrs	1G - 2H
Tue 27.06.06	55	Dortmund	17:00 hrs	1F - 2E
	56	Hanover	21:00 hrs	1H - 2G

### Quarter-finals

Fri 30.06.06	57	Berlin	17:00 hrs	W49 - W50
	58	Hamburg	21:00 hrs	W53 - W54
Sat 01.07.06	59	Geisenkirchen	17:00 hrs	W51 - W52
	60	Frankfurt	21:00 hrs	W55 - W56

### Semi-finals

Tue 04.07.06	61	Dortmund	21:00 hrs	W57 - W58
Wed 05.07.06	62	Munich	21:00 hrs	W59 - W60

### 3rd Place

Sat 08.07.06	63	Stuttgart	21:00 hrs	L61 - L62
--------------	----	-----------	-----------	-----------

### Final

Sun 09.07.06	64	Berlin	20:00 hrs	W61 - W62
--------------	----	--------	-----------	-----------

附錄二. 2006 世界盃「德國」對「哥斯大黎加」比賽報告足球  
賽單場比賽報告.



FIFA World Cup™  
Statistics



Group A

Germany - Costa Rica

4:2 (2:1)

Match Date Venue / Stadium / Country Time Att  
1 9 JUN 2006 Munich / FIFA World Cup Stadium / GER 18:00 66,000 sold out

Germany (GER)	Statistics	Costa Rica (CRC)
21	Shots	4
10	Shots on Goal	2
11	Fouls	15
7	Corner Kicks	3
1	Direct Free Kicks to Goal	0
0	Indirect Free Kicks to Goal	0
0	Penalty Kicks	0
3	Offsides	3
0	Own Goals	0
0	Cautions	1
0	Expulsions	0
63%	Ball Possession	37%
34	Actual Playing Time	20

Man of the Match: 11, KLOSE Miroslav (GER)

Germany (GER)											
#	Name	Pos	Min	GF	GA	SG/S	PK	Fouls		Y	R
								FC	FS		
1	LEHMANN Jens	GK	90		2						
3	FRIEDRICH Arne	DF	90			0/2		1			
7	SCHWEINSTEIGER Bastian	MF	90			1/1		1	3		
8	FRINGS Torsten	MF	90	1		1/2		2	2		
11	KLOSE Miroslav	FW	78	2		3/4		1	1		
16	LAHM Philipp	DF	90	1		2/2		1			
17	MERTESACKER Per	DF	90			1/1		1	1		
18	BOROWSKI Tim	MF	71			1/1		1	1		
19	SCHNEIDER Bernd (C)	MF	89			0/2					
20	PODOLSKI Lukas	FW	90			1/5		1	2		
21	METZELDER Christoph	DF	90			0/1		1			
<b>Substitutes</b>											
5	KEHL Sebastian	MF	19								
10	NEUVILLE Oliver	FW	12					1	2		
22	ODONKOR David	MF	1								
<b>Totals</b>				4	2	10/21		11	12		

Costa Rica (CRC)											
#	Name	Pos	Min	GF	GA	SG/S	PK	Fouls		Y	R
								FC	FS		
18	PORRAS Jose	GK	90		4						
3	MARIN Luis (C)	DF	90					1	1		
4	UMANA Michael	DF	90					2	1		
5	MARTINEZ Gilberto	DF	65								
6	FONSECA Danny	MF	90			0/1		1	2	1	
8	SOLIS Mauricio	MF	77					1	2		
9	WANCHOPE Paulo	FW	90	2		2/2		3	1		
10	CENTENO Walter	MF	90					1	1		
11	GOMEZ Ronald	FW	89					1	1		
12	GONZALEZ Leonardo	DF	90					4	1		
20	SEQUEIRA Douglas	MF	90					1			
<b>Substitutes</b>											
2	DRUMMOND Jervis	DF	25						1		
7	BOLANOS Cristian	MF	13			0/1					
14	AZOFEIFA Randall	MF	1								
<b>Totals</b>				2	4	2/4		15	11	1	

AET: After extra time    PSO: Penalty shoot-out    (C): Captain    Pos: Position    GK: Goalkeeper  
 D: Defender    M: Midfielder    F: Forward    Min: Minutes Played    GF: Goals For  
 R: Red Cards    GA: Goals Against    PK: Penalty Kicks (Goals/Shots)    FC: Fouls Committed    FS: Fouls Suffered  
 Y: Yellow Cards    SG/S: Shots on Goal/Shots

SAT 10 JUN 2006, 13:49 - Version 2

Page 2/2



(from <http://fifaworldcup.yahoo.com>)

## 附錄三. 2006 世界盃足球賽比賽結果.

### 第一階段比賽結果(32 強賽).

Group A					
Match	Date	Venue	Teams	Time/Score	
1	9-Jun-2006	Munich	GER: CRC	4:2 (2:1)	<a href="#">Photos</a>   <a href="#">Video</a>   <a href="#">More</a>
2	9-Jun-2006	Gelsenkirchen	POL: ECU	0:2 (0:1)	<a href="#">Photos</a>   <a href="#">Video</a>   <a href="#">More</a>
17	14-Jun-2006	Dortmund	GER: POL	1:0 (0:0)	<a href="#">Photos</a>   <a href="#">Video</a>   <a href="#">More</a>
18	15-Jun-2006	Hamburg	ECU: CRC	3:0 (1:0)	<a href="#">Photos</a>   <a href="#">Video</a>   <a href="#">More</a>
33	20-Jun-2006	Berlin	ECU: GER	0:3 (0:2)	<a href="#">Photos</a>   <a href="#">Video</a>   <a href="#">More</a>
34	20-Jun-2006	Hanover	CRC: POL	1:2 (1:1)	<a href="#">Photos</a>   <a href="#">Video</a>   <a href="#">More</a>

Group B					
Match	Date	Venue	Teams	Time/Score	
3	10-Jun-2006	Frankfurt	ENG: PAR	1:0 (1:0)	<a href="#">Photos</a>   <a href="#">Video</a>   <a href="#">More</a>
4	10-Jun-2006	Dortmund	TRI: SWE	0:0	<a href="#">Photos</a>   <a href="#">Video</a>   <a href="#">More</a>
19	15-Jun-2006	Nuremberg	ENG: TRI	2:0 (0:0)	<a href="#">Photos</a>   <a href="#">Video</a>   <a href="#">More</a>
20	15-Jun-2006	Berlin	SWE: PAR	1:0 (0:0)	<a href="#">Photos</a>   <a href="#">Video</a>   <a href="#">More</a>
35	20-Jun-2006	Cologne	SWE: ENG	2:2 (0:1)	<a href="#">Photos</a>   <a href="#">Video</a>   <a href="#">More</a>
36	20-Jun-2006	Kaiserslautern	PAR: TRI	2:0 (1:0)	<a href="#">Photos</a>   <a href="#">Video</a>   <a href="#">More</a>

Group C					
Match	Date	Venue	Teams	Time/Score	
5	10-Jun-2006	Hamburg	ARG: CIV	2:1 (2:0)	<a href="#">Photos</a>   <a href="#">Video</a>   <a href="#">More</a>
6	11-Jun-2006	Leipzig	SCG: NED	0:1 (0:1)	<a href="#">Photos</a>   <a href="#">Video</a>   <a href="#">More</a>
21	16-Jun-2006	Gelsenkirchen	ARG: SCG	6:0 (3:0)	<a href="#">Photos</a>   <a href="#">Video</a>   <a href="#">More</a>
22	16-Jun-2006	Stuttgart	NED: CIV	2:1 (2:1)	<a href="#">Photos</a>   <a href="#">Video</a>   <a href="#">More</a>
37	21-Jun-2006	Frankfurt	NED: ARG	0:0	<a href="#">Photos</a>   <a href="#">Video</a>   <a href="#">More</a>
38	21-Jun-2006	Munich	CIV: SCG	3:2 (1:2)	<a href="#">Photos</a>   <a href="#">Video</a>   <a href="#">More</a>

Group D					
Match	Date	Venue	Teams	Time/Score	
7	11-Jun-2006	Nuremberg	MEX: IRN	3:1 (1:1)	<a href="#">Photos</a>   <a href="#">Video</a>   <a href="#">More</a>
8	11-Jun-2006	Cologne	ANG: POR	0:1 (0:1)	<a href="#">Photos</a>   <a href="#">Video</a>   <a href="#">More</a>
23	16-Jun-2006	Hanover	MEX: ANG	0:0	<a href="#">Photos</a>   <a href="#">Video</a>   <a href="#">More</a>
24	17-Jun-2006	Frankfurt	POR: IRN	2:0 (0:0)	<a href="#">Photos</a>   <a href="#">Video</a>   <a href="#">More</a>
39	21-Jun-2006	Gelsenkirchen	POR: MEX	2:1 (2:1)	<a href="#">Photos</a>   <a href="#">Video</a>   <a href="#">More</a>
40	21-Jun-2006	Leipzig	IRN: ANG	1:1 (0:0)	<a href="#">Photos</a>   <a href="#">Video</a>   <a href="#">More</a>

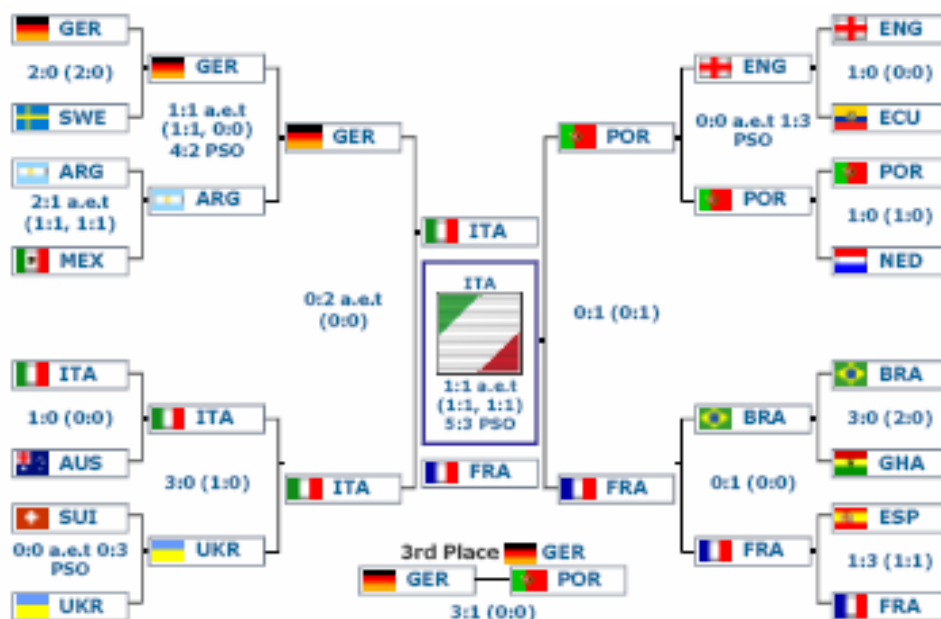
Group E					
Match	Date	Venue	Teams	Time/Score	
9	12-Jun-2006	Hanover	ITA: GHA	2:0 (1:0)	<a href="#">Photos</a>   <a href="#">Video</a>   <a href="#">More</a>
10	12-Jun-2006	Gelsenkirchen	USA: CZE	0:3 (0:2)	<a href="#">Photos</a>   <a href="#">Video</a>   <a href="#">More</a>
25	17-Jun-2006	Kaiserslautern	ITA: USA	1:1 (1:1)	<a href="#">Photos</a>   <a href="#">Video</a>   <a href="#">More</a>
26	17-Jun-2006	Cologne	CZE: GHA	0:2 (0:1)	<a href="#">Photos</a>   <a href="#">Video</a>   <a href="#">More</a>
41	22-Jun-2006	Hamburg	CZE: ITA	0:2 (0:1)	<a href="#">Photos</a>   <a href="#">Video</a>   <a href="#">More</a>
42	22-Jun-2006	Nuremberg	GHA: USA	2:1 (2:1)	<a href="#">Photos</a>   <a href="#">Video</a>   <a href="#">More</a>

Group F					
Match	Date	Venue	Teams	Time/Score	
11	13-Jun-2006	Berlin	BRA: CRO	1:0 (1:0)	<a href="#">Photos</a>   <a href="#">Video</a>   <a href="#">More</a>
12	12-Jun-2006	Kaiserslautern	AUS: JPN	3:1 (0:1)	<a href="#">Photos</a>   <a href="#">Video</a>   <a href="#">More</a>
27	18-Jun-2006	Munich	BRA: AUS	2:0 (0:0)	<a href="#">Photos</a>   <a href="#">Video</a>   <a href="#">More</a>
28	18-Jun-2006	Nuremberg	JPN: CRO	0:0	<a href="#">Photos</a>   <a href="#">Video</a>   <a href="#">More</a>
43	22-Jun-2006	Dortmund	JPN: BRA	1:4 (1:1)	<a href="#">Photos</a>   <a href="#">Video</a>   <a href="#">More</a>
44	22-Jun-2006	Stuttgart	CRO: AUS	2:2 (1:1)	<a href="#">Photos</a>   <a href="#">Video</a>   <a href="#">More</a>

Group G					
Match	Date	Venue	Teams	Time/Score	
13	13-Jun-2006	Stuttgart	FRA: SUI	0:0	<a href="#">Photos</a>   <a href="#">Video</a>   <a href="#">More</a>
14	13-Jun-2006	Frankfurt	KOR: TOG	2:1 (0:1)	<a href="#">Photos</a>   <a href="#">Video</a>   <a href="#">More</a>
29	18-Jun-2006	Leipzig	FRA: KOR	1:1 (1:0)	<a href="#">Photos</a>   <a href="#">Video</a>   <a href="#">More</a>
30	19-Jun-2006	Dortmund	TOG: SUI	0:2 (0:1)	<a href="#">Photos</a>   <a href="#">Video</a>   <a href="#">More</a>
45	23-Jun-2006	Cologne	TOG: FRA	0:2 (0:0)	<a href="#">Photos</a>   <a href="#">Video</a>   <a href="#">More</a>
46	23-Jun-2006	Hanover	SUI: KOR	2:0 (1:0)	<a href="#">Photos</a>   <a href="#">Video</a>   <a href="#">More</a>

Group H					
Match	Date	Venue	Teams	Time/Score	
15	14-Jun-2006	Leipzig	ESP: UKR	4:0 (2:0)	<a href="#">Photos</a>   <a href="#">Video</a>   <a href="#">More</a>
16	14-Jun-2006	Munich	TUN: KSA	2:2 (1:0)	<a href="#">Photos</a>   <a href="#">Video</a>   <a href="#">More</a>
31	19-Jun-2006	Stuttgart	ESP: TUN	3:1 (0:1)	<a href="#">Photos</a>   <a href="#">Video</a>   <a href="#">More</a>
32	19-Jun-2006	Hamburg	KSA: UKR	0:4 (0:2)	<a href="#">Photos</a>   <a href="#">Video</a>   <a href="#">More</a>
47	23-Jun-2006	Kaiserslautern	KSA: ESP	0:1 (0:1)	<a href="#">Photos</a>   <a href="#">Video</a>   <a href="#">More</a>
48	23-Jun-2006	Berlin	UKR: TUN	1:0 (0:0)	<a href="#">Photos</a>   <a href="#">Video</a>   <a href="#">More</a>

第二階段比賽結果(含 16 強賽、1/4 決賽、1/2 決賽及決賽).



Round of 16

Match	Date	Venue	Teams	Time/Score	Photos	Video	More
49	24 June 2006	Munich	GER:SWE	2:0 (2:0)	<a href="#">Photos</a>	<a href="#">Video</a>	<a href="#">More</a>
50	24 June 2006	Leipzig	ARG:MEX	2:1 a.e.t (1:1, 1:1)	<a href="#">Photos</a>	<a href="#">Video</a>	<a href="#">More</a>
51	25 June 2006	Stuttgart	ENG:ECU	1:0 (0:0)	<a href="#">Photos</a>	<a href="#">Video</a>	<a href="#">More</a>
52	25 June 2006	Nuremberg	POR:NED	1:0 (1:0)	<a href="#">Photos</a>	<a href="#">Video</a>	<a href="#">More</a>
53	26 June 2006	Kaiserslautern	ITA:AUS	1:0 (0:0)	<a href="#">Photos</a>	<a href="#">Video</a>	<a href="#">More</a>
54	26 June 2006	Cologne	SUI:UKR	0:0 a.e.t 0:3 PSO	<a href="#">Photos</a>	<a href="#">Video</a>	<a href="#">More</a>
55	27 June 2006	Dortmund	BRA:GHA	3:0 (2:0)	<a href="#">Photos</a>	<a href="#">Video</a>	<a href="#">More</a>
56	27 June 2006	Hanover	ESP:FRA	1:3 (1:1)	<a href="#">Photos</a>	<a href="#">Video</a>	<a href="#">More</a>

Quarter-finals

Match	Date	Venue	Teams	Time/Score	Photos	Video	More
57	30 June 2006	Berlin	GER:ARG	1:1 a.e.t (1:1, 0:0) 4:2 PSO	<a href="#">Photos</a>	<a href="#">Video</a>	<a href="#">More</a>
58	30 June 2006	Hamburg	ITA:UKR	3:0 (1:0)	<a href="#">Photos</a>	<a href="#">Video</a>	<a href="#">More</a>
59	1 July 2006	Gelsenkirchen	ENG:POR	0:0 a.e.t 1:3 PSO	<a href="#">Photos</a>	<a href="#">Video</a>	<a href="#">More</a>
60	1 July 2006	Frankfurt	BRA:FRA	0:1 (0:0)	<a href="#">Photos</a>	<a href="#">Video</a>	<a href="#">More</a>

Semi-Finals

Match	Date	Venue	Teams	Time/Score	Photos	Video	More
61	4 July 2006	Dortmund	GER:ITA	0:2 a.e.t (0:0)	<a href="#">Photos</a>	<a href="#">Video</a>	<a href="#">More</a>
62	5 July 2006	Munich	POR:FRA	0:1 (0:1)	<a href="#">Photos</a>	<a href="#">Video</a>	<a href="#">More</a>

Third Place

Match	Date	Venue	Teams	Time/Score	Photos	Video	More
63	8 July 2006	Stuttgart	GER:POR	3:1 (0:0)	<a href="#">Photos</a>	<a href="#">Video</a>	<a href="#">More</a>

Final

Match	Date	Venue	Teams	Time/Score	Photos	Video	More
64	9 July 2006	Berlin	ITA:FRA	1:1 a.e.t (1:1, 1:1) 5:3 PSO	<a href="#">Photos</a>	<a href="#">Video</a>	<a href="#">More</a>