

第一章 緒論

1-1. 研究背景與動機：

資訊科技的快速發展，加速了硬體設備的汰換與升級。資料儲存設備(Storage)與資料存取的技术與方法,也不斷地推陳出新。90 年代以來，企業的經營範圍逐漸由地區性轉變為全球性，而經營策略也逐漸由提昇品質轉向為速度的競爭。藉由發達的網路通訊技術，將來自各地區之資料匯入大型的「資料市集」(data mart)或是「資料倉儲」(data warehouse)之內。此類大型的資料庫，不但可以儲存完整的歷史資料，並可利用「資料探勘」(data mining)技術或是線上分析處理技術(On-Line Analytical Processing ,OLAP)，在龐大的資料中，找出資料模式(data pattern)彼此間的隱性關聯，分析出有利於企業經營之資訊，藉此作為制定未來營運策略方針之依據。

目前在資料探勘的應用領域中，大部分的研究係針對在大量資料庫內，找出各個獨立資料彼此間的關聯法則，藉以分析點狀分散式資料中的知識擷取。但在擷取大量資料的過程當中，常會發現資料彼此之間具有時間延續的依存關係，也就是呈現線狀分布的時間序列(time series)資料[Donald J. Berndt and James Clifford,1995]。時間序列資料是指在一段特定期間，每個時間點將會對應到一個數值。這個因時間改變而變化的數值，就應用層面來看，可以為：生產量、銷售量、

股價、降雨量等。與時間相關的資料探勘領域中，對未來趨勢預測的技術，已逐漸成為企業經營模式中，不可缺少的一環。學者 Snodgrass 曾說：「找尋因時間而變化的行為模式，其彼此間的關聯性，將是未來時間性資料管理的重心之一」。[Davood Rafiei,1999]

預測(prediction)是資料探勘中的一種模式。根據歷史資料，透過特定的變數值訓練組資料，建立預測模型。並輸入最新的資料，將可求得對未來趨勢變化的預測值[Michael J.A. Berry and Gordon S. Linoff,1997]。針對長時間性的資料庫(temporal database)作特定變異因子的預測分析，歷史資料的筆數多寡與資料在時間上的連續性，均會影響預測的結果。

以台灣地區蔬果內需交易市場為例，無論是產地市場或批發市場，每日蔬果的交易量都極為龐大。以台灣每人每年平均蔬果消費量為例，如[表 1]所示，蔬菜由民國 50 年的 57.18 公斤提高至民國 87 年的 104.93 公斤，增加一倍有餘。水果更由民國 50 年的 19.86 公斤提高至民國 87 年的 135.53 公斤，增加 6.8 倍。由此可充分顯示，台灣蔬果內需市場的交易價格與數量，不但符合大量資料的型態，且因每日皆有交易明細資料，亦符合時間性資料庫的特性。

就特定的變異因子的選取上，以台灣地區生產的蔬菜為例，在季節的分布上，產量有冬季多而夏季少的狀況。由於一年內各季的供應量不同，價格亦呈現明顯的負相關，如[圖 1-1]。主要產地的蔬菜生產

量，極易受到氣象因素的影響[楊新輝,1999]，而導致批發市場交易價格的波動。常態性的氣候因素有：溫度、溼度、蒸發皿蒸發量、日照、降雨量等[洪英琳,1998]及非常態性的氣候因素，例如：颱風的侵襲[楊新輝,1999]。

因此，本研究將以 1998/01/01-2000/04/30 兩年又四個月的每日蔬菜批發交易之價量作為時間序列的歷史資料庫，探討如何利用「資料探勘」技術，發掘特定變因對時間性資料庫中歷史資料的影響，並作為未來趨勢預測之依據。

1-2 研究目的：

本研究擬針對台灣地區主要的蔬菜產地-彰化縣溪湖鎮與雲林縣西螺鎮兩地交易市場[宋振裕,1995]的每日交易狀況與當地的氣候變因，作一整體性之探討。就如何利用「監督式知識發掘」(supervised knowledge discovery)的程序與步驟，發掘資料彼此間有意義的規則，根據既有的歷史資料，和對過去事件的解釋，作為未來預測的基礎 [Michael J.A. Berry and Gordon S. Linoff, 1997]。

並以類神經網路(artificial neural network;簡稱 ANN)中的倒傳遞網路模式(back propagation network; 簡稱 BPN), 作為預測方法，建立氣象因素對台灣地區主要蔬菜產地之蔬菜栽種的預測模式，以期完成下述目的：

1. 探討利用「監督式知識發掘」發掘時間性資料庫中資料的關聯特性
2. 了解台灣地區主要批發市場蔬菜交易的價量關係。
3. 探討氣象因素對產地蔬菜栽種的影響。
4. 運用 BPN 類神經網路模式分析的方法，發掘氣象因素與蔬菜生產量彼此間的關係，並嘗試建立產地蔬菜季節性栽種之預測模式，以提供農政單位輔導農民作計劃性栽種之參考依據。

1-3 研究範圍：

基於研究動機所述，本研究範圍概述如下：

- 一、 在六種資料探勘技術中[Michael J.A. Berry and Gordon S. Linoff, 1997]，本研究係採取「預測」(Prediction)方式。在時間性資料庫中，可透過歷史資料用來建立模型，以檢視最新輸入資料值，可推測未來的變化的預測。
- 二、 本研究所使用的時間性資料庫，經分析後發現，大台北地區之蔬果消費量占台灣地區總消費量之四成，為主要的批發市場。而在蔬菜方面，以甘藍、包心白菜和青江菜為成交量的前三名。根據本研究整理發現，就西螺與溪湖產地批發市場的交易紀錄觀察，一年十二個月皆有甘藍菜(初秋品種)交易紀錄，符合時間序列特性，故特別以二年度(1998-1999)甘藍菜產地交易量作為類神經網路訓練的基本樣本。
- 三、 變異因素資料來源為中央氣象局所提供的逐日地面氣象資料(A、B表)，記錄時間是 1998/01/01-1999/12/31。

1-4 研究流程：

本研究之流程為：先確定研究主題，藉由歷史文獻的回顧與探討，決定使用「監督式知識發掘」(supervised knowledge discovery)中的 BPN 類神經網路演算模式作為研究方法，繼而開始收集台灣地區主要蔬菜交易批發市場之日交易量與交易價格，及影響蔬菜生長的變異因子-氣象資料。資料來源為台灣省政府農林廳所出版的農業統計年報 [省政府農林廳,1998]，行政院農業委員會農產品行情交易網站和農業產銷班資訊服務網所提供之資料，本研究逐年蒐集整理。在變異因子資料部分，資料來源為中央氣象局所提供的逐日地面氣象資料(A、B表)。利用 Microsoft Excel 2000 整理各項基本資料，再使用 MATLAB 建置 BPN 類神經網路演算模式，以進行測試。根據測試結果分析氣象因素對產地蔬菜產量的關係，以作為未來蔬菜栽種之預測模式。本研究流程圖如[圖 1-2]所示。

1-5 本文架構：

本研究共分五章。架構如下：

第一章 緒論：說明研究背景與動機、研究目的、研究範圍與限制、研究流程與本文架構。

第二章 文獻探討：介紹「資料探勘」和「知識發掘」的定義與兩者之間的差異。就「監督式知識發掘」的定義與方法作文獻回顧。並介紹台灣地區蔬菜批發交易市場的價量關係和氣象因素對產地蔬菜生長的影響。最後根據文獻探討類神經網路。

第三章 研究方法與架構：針對 BPN 類神經網路簡述其理論架構和演算模式，及使用上應注意的事項。

第四章 實驗設計與結果分析：本章將說明本研究之預測方法，變異因數的選取，及衡量預測誤差標準。並說明如何建立 BPN 類神經網路模型，以實證分析氣象因素對台灣地區主要蔬菜產地的產量影響。

第五章 結論與未來研究方向：針對實證後所得到的研究結果加以探討比較並提出未來的研究方向之建議。

表 1-1、台灣地區每人每年糧食可供消費量(資料期間:民國 50 年-民國 87 年)

* 單位：公斤

民國	穀類			蔬菜	水果	肉類	魚介類
	總計	白米	麵粉				
50	165.03	136.78	24.57	57.18	19.86	15.60	25.33
51	156.81	132.10	22.04	56.16	21.62	16.04	26.14
52	156.77	134.36	18.47	60.28	20.74	17.90	27.27
53	150.84	129.87	17.17	56.58	17.81	18.45	28.20
54	157.34	132.85	22.33	56.81	21.01	19.21	27.74
55	156.33	137.42	16.58	52.67	26.18	22.91	28.84
56	157.80	141.47	13.43	52.49	34.11	26.26	28.69
57	162.59	139.93	20.37	67.56	52.65	27.08	29.69
58	166.67	138.74	24.88	74.97	38.45	23.32	30.32
59	164.11	134.45	25.42	84.83	45.83	25.25	34.18
60	164.63	134.28	25.51	91.27	44.96	26.43	34.34
61	165.78	133.52	27.10	91.17	41.75	27.34	35.28
62	161.58	129.84	26.96	92.98	50.23	28.71	36.96
63	167.81	134.15	28.72	98.93	62.93	27.45	34.25
64	162.11	130.39	24.21	109.79	55.03	26.98	35.56
65	156.25	128.12	20.78	118.39	62.14	31.64	35.27
66	156.63	125.06	22.58	122.37	57.41	35.26	35.05
67	146.46	113.99	23.92	114.93	54.32	36.12	36.48
68	142.39	105.27	23.81	127.50	66.62	40.26	38.09
69	134.08	100.82	23.62	129.58	70.16	42.62	38.74
70	128.76	96.54	23.38	115.60	80.51	42.99	35.79

民國	穀類			蔬菜	水果	肉類	魚介類
	總計	白米	麵粉				
71	127.56	93.07	23.29	118.21	71.54	46.36	35.16
72	121.95	89.33	24.21	116.61	68.42	50.02	33.09
73	114.20	84.40	23.27	109.63	105.85	52.77	35.55
74	110.63	80.19	24.96	103.41	111.50	55.75	35.12
75	111.54	76.46	27.50	93.14	105.13	56.47	35.63
76	110.81	73.33	29.06	102.93	126.30	58.89	42.27
77	109.80	70.14	31.52	98.53	135.49	59.15	41.35
78	105.78	68.27	29.71	98.29	135.70	61.34	45.49
79	102.48	65.94	28.65	93.33	131.50	62.89	47.48
80	99.80	62.50	28.49	94.73	138.69	64.51	39.72
81	100.67	62.23	29.00	97.46	129.82	67.17	43.95
82	99.61	60.70	28.93	98.06	144.58	70.33	47.46
83	102.02	59.89	31.57	93.33	136.54	72.39	38.37
84	100.66	59.10	31.88	101.86	137.42	72.76	38.35
85	97.42	58.84	29.57	108.85	138.83	75.04	38.40
86	96.20	58.40	30.27	108.27	150.08	77.31	42.35
87	95.83	56.74	31.90	104.93	135.53	78.77	40.43

資料來源：本研究整理

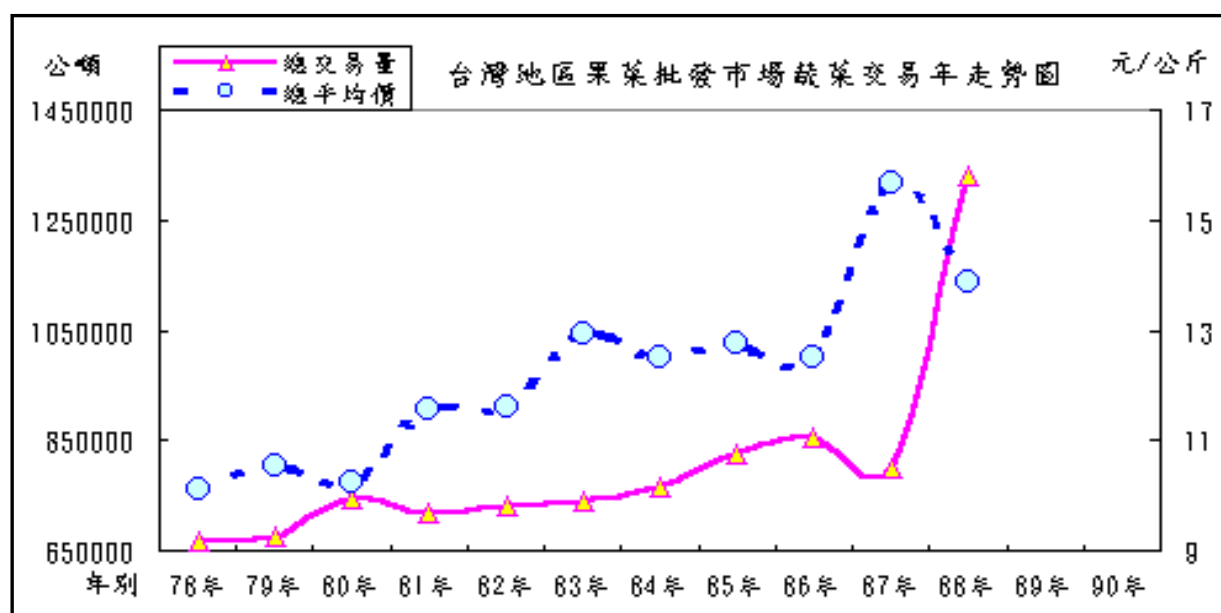


圖 1-1、台灣地區果菜批發市場 78-89 年度每月蔬菜交易量、價變化走勢圖

資料來源:行政院農業委員會農產品行情交易網站

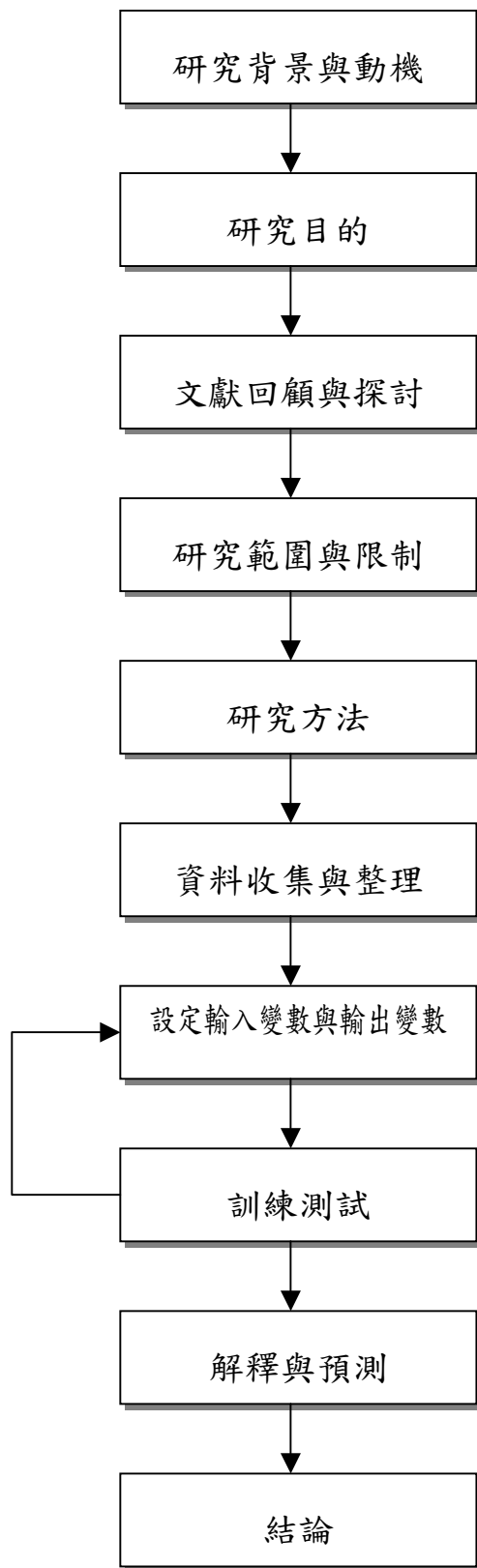


圖 1-2、研究流程圖