

基於「知識管理」檢討資訊管理系統之發展

張庭魁

雲林科技大學資管系

蔡鴻旭

虎尾科技大學資管系

摘要

雲林縣政府於 2003 年開始陸續推動知識管理系統，提供文件儲存與檢索、電子表單、公告系統等功能。雖然資訊系統大部份都會留存使用行為紀錄，然而在雲林縣政府推動知識管理系統的過程中，使用者或相關系統的分析研究卻付之闕如，因此本研究整合 Proxy 紀錄檔與員工資料檔，以發掘知識，提供擬訂相關資訊系統發展政策。

本研究的主要貢獻在於首次整合分析層級程序法 (Analytic Hierarchy Process) 及資料探勘 (Data Mining) 技術，將其應用在政府相關資訊系統的評估。本研究利用分析層級程序法以取得各系統之貢獻權重，並利用相關權重，結合資料探勘方法，以評估雲林縣政府內使用知識管理系統的員工。資料探勘所採用之演算法分別為 K-means 演算法及決策樹 (Decision Tree) 演算法，所分析之系統重要性及員工資料結果，將可提供雲林縣政府擬定新的資訊系統發展策略。

關鍵字：分析層級程序法、顧客生命週期價值、資料探勘、知識管理、雲林縣政府

[收稿]2008/6/03; [初審] 2008/8/17; [接受刊登] 2008/12/31

通訊作者：張庭魁(e-mail: changtk@yuntech.edu.tw)

壹、前言

在知識管理被提出多年之後，很少文獻能以量化方式並以內部觀點來討論知識管理，因此本研究以知識管理環境為前提，並以資訊系統為研究範圍，討論相關資訊系統及內部使用者的背景，以做為制定資訊系統發展政策的建議。下列將分別說明本研究的方法及資料內容。

首先，在研究方法部分，本研究的貢獻在於有效整合運用「分析層級程序法」與「資料探勘」。以往「顧客生命週期價值」(Customer Lifetime Value, CLV)中常用 RFM (Recency, Frequency, Monetary)等三項變數來評估對企業有價值的顧客，在其他文獻中較常看見在 M 變數(Monetary)中是以貨幣金額等計算(Shih, 2003; Liu, 2005; Xu, 2003; Marcus, 1998; Rust, 2005; Buckinx, 2005; King, 2007)；而在政府機關的相關活動中，大部份卻是無法以金額計算，與非營利性組織情況類似，故不適合以貨幣金額做為 M 變數的值。

King(2007)曾提到一種模式為 RFC (Recency, Frequency, Cost)，該文獻中主張在政府機構中，以 RFC 取代 RFM 模式；此文獻主要是主張在 Cost 變數的值，以民眾使用服務的直接財務成本予以計算，或是間接以民眾使用該服務的有效期間所耗費的質量或是民眾使用該服務後造成的那些影響，此文獻的主要訴求在「服務成本」，這與本研究稍有差異。

本研究以「分析層級程序法」讓評估者評估系統的權重，並據以評估使用者對知識管理相關系統有多少「貢獻」，因為本研究主要是發現具價值性的使用者，而非最具耗費服務成本的使用者，所以採用「分析層級程序法」評估權重後，以該權重當做 M 變數的值更具有意義。

而本研究的次要貢獻，是以「分析層級程序法」對系統進行評估，有助於決策者瞭解究竟那些系統對組織發展知識管理是有助益的，那些提供的助益較少。最後，本研究也基於以往「顧客關係管理」(Customer Relationship Management, CRM)及「顧客生命週期價值」的

觀念，首次針對政府機關的內部使用者，提出以忠誠度及價值性為基礎的分類方式，並針對不同類別使用者提出不同的決策建議。

而本研究的資料內容部分，所討論的知識管理環境之範疇，係採取 Tiwana(2002)對知識管理環境的定義，研究資料的範圍限定在資訊相關系統中。Tiwana 將知識管理環境區分為七層，本研究將依照這七層的定義，在雲林縣政府現有相關的資訊平台中予以區分成七種，再加以分析。

而分析將分成兩種不同程序進行，第一種是不分層的狀況下去分析，第二種是以分層的方式去分析。分層分析的用意在於將不同貢獻程度的系統分開討論；例如「合作層」中相關系統的貢獻度比「應用層」中相關系統的貢獻度低，自然不應混合一起分析其價值性，故第二種分析方式，將所有系統區分為「合作及智慧層」、「應用層」、「傳輸層」，再分別予以分析。

因為雲林縣政府所擁有之資訊系統並未完全涵蓋在 Tiwana 的知識管理七層定義內，僅能找出有關「合作層」、「應用層」及「傳輸層」之系統，因為本研究主要聚焦在資訊系統的檢討及使用者行為，只是套用知識管理的分層方式區分並予以賦權，故不影響研究結果。

貳、問題描述

因為雲林縣政府推動知識管理已有一段時間，但一直沒有相關資料可以進行知識管理環境的分析，如果再想進行使用者的背景分析更是困難，因為會牽涉到不同的單位及不同的資訊系統。因此本研究嘗試瞭解：

1. 使用者有那些？這些使用者的背景為何？針對使用者背景，如何制定相關決策？
2. 新進員工進入縣府時，是否能預測其背景，以加快培訓程序？
3. 已建置之知識管理系統效益如何？

而本研究的結果，期望能提供雲林縣政府那些貢獻？以下則是本

研究三項主要效益：

- 1.提供擬定組織策略之參考。
- 2.提供訓練課程規劃之建議。
- 3.提供資訊管理系統之檢討及發展策略。

參、相關文獻

本節將說明知識管理的定義，另外也提出與「分析層級程序法」及「顧客生命週期價值」的相關文獻，最後將說明要進行資料探勘之前，須進行的資料預處理有那些。

甲、知識管理

我們知道「知識管理」的解釋眾說紛紜，有如戴文坡 & 普賽克(2001)對知識管理採取較廣的定義，也有如 Tiwana(2002)對知識管理採取範圍較小的定義。而本研究採用 Tiwana 對知識管理環境的分層方式予以區分系統。Tiwana 對於知識管理平台有細膩的定義及分類，非常適合用來評估以資訊系統為基礎的知識管理平台。

乙、「分析層級程序法」與「顧客生命週期價值」的應用

為了解決複雜的決策問題，美國 Saaty 於一九七一年提出「分析層級程序法」。現今「分析層級程序法」已大量應用在各領域中做為決策等相關應用，如公共工程、醫學、商業、電力管理、國防、油管輸送等等。(Shih, 2003; Liu, 2005; Greiner, 2003; Wang, 2005; Dey, 2004; Agalgaonkar, 2006; Sloane, 2004)。

另外，「顧客關係管理」在公領域也逐漸被拿出來討論，這也導致產生另外一種「公民關係管理」(Citizen Relationship Management, CRM)，那麼在政府部門談「顧客關係管理」究竟有何助益？Themistocleous(2005)詳述幾種利益，分別為降低成本、支援資訊分享、提供較佳的服務、改進服務項目的功能性、達成服務個人化、支

援決策程序、增進服務相關知識、增加公民的滿意度等等。本研究所使用 RFM 是評估「顧客生命週期價值」的重要測量工具之一，根據 Bult 及 Wansbeek(1995)的定義如下：

- **R**: 在一定期間內，最接近的一次交易時間。
- **F**: 在一定期間內，交易總次數。
- **M**: 在一定期間內，交易的總金額。

而在本研究中，首次提出以「分析層級程序法」方式，由相關領域的專業主管及員工共同決策，決定各系統的價值權重，取代傳統以貨幣價金的計算方式。再將使用紀錄結合 RFM 三項變數以資料探勘的分群方式分群，再根據這些分群資料分別研訂決策。

在顧客關係管理的討論範圍中，除了外部顧客之外，其實還有一個很重要的內部顧客常為我們所忽略，因此如何促進員工的關係管理，進而提昇組織運作效率，尤其在資訊系統的應用上，也是本研究所關心的重點。

丙、資料預處理及資料探勘

在資料探勘之前，最耗時的工作莫過於資料的預處理，所謂資料的預處理包括有「資料清理」、「資料整合」、「資料轉換」、「資料化約」等等。而在一些文獻則是將資料預處理分成「資料清理」(Data Cleaning)，「使用者之識別」(User Identification)，「連結會談之識別」(Session Identification) 及「路徑資料補正」(Path Complement) 等四個程序，詳見文獻(Srivastava, 2000; Yuan, 2003; Huiying, 2004)，流程請見圖 1。

而資料探勘主要功能可以區分為 Classification、Regression、Time Series、Clustering、Summarization、Association Rules、Sequence Discovery 等(Dunham, 2002)。本研究先 K-means 演算法將 RFM 變數相近者予以分群，再使用決策樹演算法來將資料分類，並預測其規則。

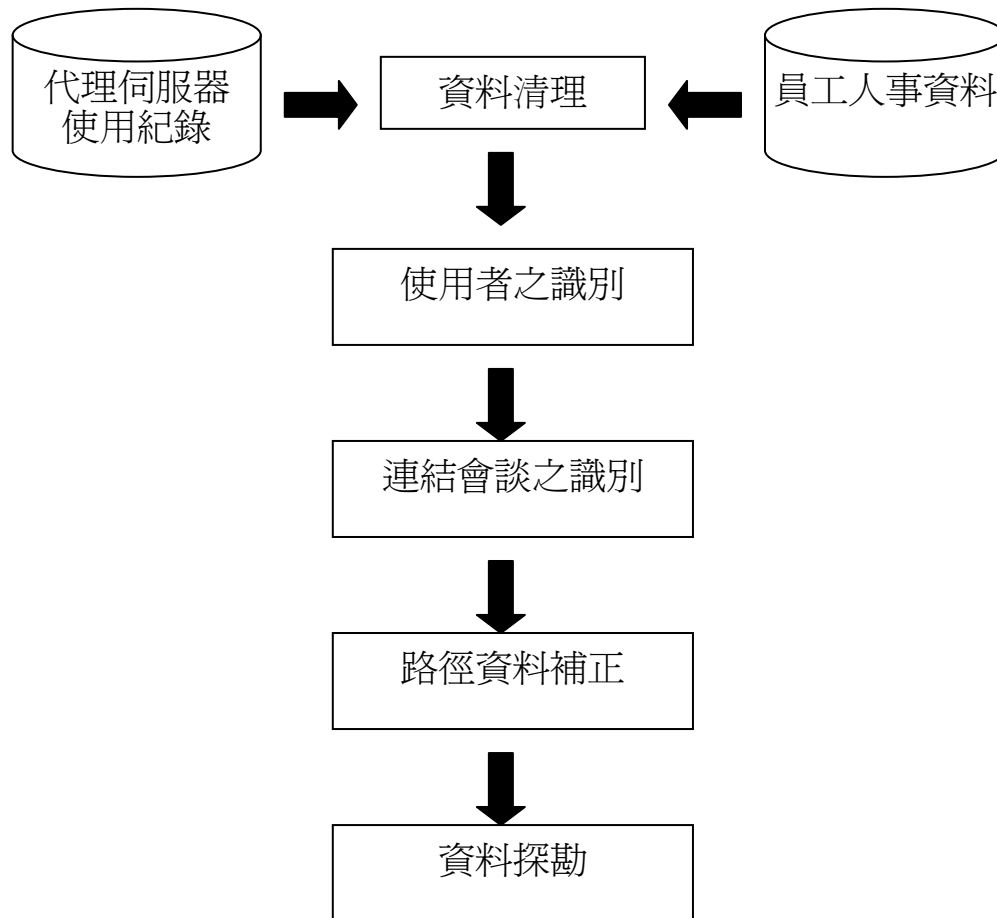


圖 1 資料預處理步驟

本研究採用決策樹演算法來將資料分類，並預測其規則。下列將說明如何在決策樹中選擇分歧的屬性，假設屬性 A 有不同的值 $\{a_1, a_2, \dots, a_v\}$ ，屬性 A 可以將 S 集合分成 v 個子集合 $\{S_1, S_2, \dots, S_v\}$ ，那麼 A 劃分子集合的 Entropy（或期望值）可由(1)求得。

$$E(A) = \sum_{j=1}^v \frac{S_{1j} + \dots + S_{mj}}{S} I(S_{1j}, \dots, S_{mj}) \quad (1)$$

而(1)中 $\frac{S_{1j} + \dots + S_{mj}}{S}$ 為第 j 個子集合的權重，即子集合的項目除以 S 集合的總數。而(1)中的 $I = (S_1, S_2, \dots, S_{mj})$ 可由(2)求得。

$$I = (S_1, S_2, \dots, S_{mj}) = - \sum_{i=1}^m p_{ij} \log_2(p_{ij}) \quad (2)$$

最後可以求得各個屬性的增益值（Gain），最高增益值的將被選

取作測試屬性，建立分枝，如(3)。

$$Gain(A) = I(S_1, S_2, \dots, S_m) - E(A) \quad (3)$$

而在資料分群部份，本研究使用 K-means 演算法，該演算法源自於 1967 年 J. B. MacQueen 所發表，詳見(4)，它具有簡易且快速的特性。

$$E = \sum_{i=1}^k \sum_{p \in C_i} |p - m_i|^2 \quad (4)$$

E 表示所有資料平方差的總和。

p 表示我們給定目標的那一點。

m_i 表示群聚 C_i 的平均值。

肆、研究方式

本節將分別介紹「研究範圍及對象」、「研究架構及進行方式」，進行方式包括有「分析層級程序法」、「資料預處理」及「資料探勘」。本章以概觀的角度，除說明雲林縣政府推動知識管理的環境背景之外，也說明本研究進行過程概要。

甲、研究範圍及對象

本研究進行當時雲林縣政府的知識管理環境已推動多年，在這之前，雲林縣政府的知識管理相關資訊系統大概是以簡易網頁及電子郵件為主。在 2003 之後，雲林縣政府擬定「2003 雲林縣政府知識管理基礎建置計畫」，在行政院研考會與行政院主計處的協助下，陸續執行建置完成相關知識管理基礎建設，在 2006 年時，雲林縣政府的知識管理環境如圖 2：

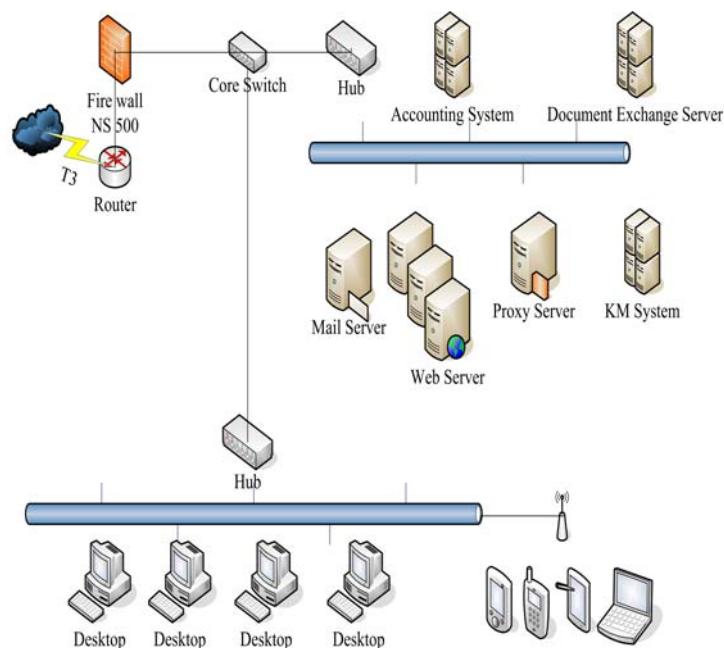


圖 2 2006 年縣府知識管理相關資訊系統架構

可以發現雲林縣政府已增加許多應用系統及協同作業系統，尤其是量及質上都有明顯增加，例如電子郵件，2003 年以前只有公文登記桌才有使用，到了 2006 年時，超過 2/3 以上的員工都已擁有電子郵件帳號。本研究將相關系統整理如表 1。

表 1 2006 年雲林縣政府主要資訊系統

系統名稱	數量	說明
Document Management Server	4	公文管理系統
IDS Policy Server	1	入侵偵測政策管理系統
IDS	1	入侵偵測系統
Microsoft SUS	1	微軟軟體更新主機
Computer Facility Monitor	1	電腦機房監控系統
Server Farm Switch	1	核心交換器
KM System	3	知識管理系統
Document Exchange Server (XML-Box)	2	公文交換主機
DNS	1	網域名稱主機
Proxy	1	代理伺服器主機
OfficeScan and IMSS	1	防毒主機

WEB Server	3	網頁主機
Mail Server	1	郵件主機
Accounting System	8	單位預算會計系統
Statistical Database Management System	2	統計資料庫

資料來源：本研究整理

因為雲林縣政府進行知識管理環境的建置已進行近三年，但一直沒有任何機會與環境針對所建置的環境進行分析，本研究希望可以透過這些員工背景資料分析，進一步瞭解雲林縣政府的知識管理環境，並據以制定政策。

首先分析的系統中有一個重要系統，就是知識管理系統，這一個系統是在 2003 年由行政院研考會補助雲林縣政府建置而成，主要功能有電子表單、文件分享、訊息公告等等。茲將該系統中較重要的幾個子系統及其網址詳列如表 2。

表 2 雲林縣政府知識管理系統之各主要子系統

各子系統	URL address
表單設計	http://eip.yunlin.gov.tw/E-Form/
表單審核	http://eip.yunlin.gov.tw/E-Sign/
差勤系統	http://eip.yunlin.gov.tw/absent/
差旅費申請	http://eip.yunlin.gov.tw/Pay/
會議室申請	http://eip.yunlin.gov.tw/EXTMeeting/
公務車申請	http://eip.yunlin.gov.tw/CarR/
財產管理	http://eip.yunlin.gov.tw/Property/
表單流程管理	http://eip.yunlin.gov.tw/FlowEngine/
網路公事包	http://eip.yunlin.gov.tw/FileCtrl/
問卷調查	http://eip.yunlin.gov.tw/Quest/
公告系統	http://eip.yunlin.gov.tw/Post/
檢索系統	http://eip.yunlin.gov.tw/SearchKM/
網路文件管理	http://eip.yunlin.gov.tw/EduInfo/
社群服務	http://eip.yunlin.gov.tw/Group/

論壇	http://eip.yunlin.gov.tw/Discuss/
管考系統	http://eip.yunlin.gov.tw/Control/
聊天室	http://eip.yunlin.gov.tw/Chat/
交辦事項	http://eip.yunlin.gov.tw/Task/

資料來源：本研究整理

除了組織內建立的相關系統之外，在外部也有不少系統其實是符合 Tiwana 所定義的知識管理相關系統，列出幾項內部員工常用的相關系統，如表 3。

表 3 網際網路上之各種知識管理相關網站

The function of web sites	URL address
論壇	http://groups.google.com.tw
	http://forum.pchome.com.tw
	http://board.yam.com
搜尋引擎	http://www.google.com.tw
	http://tw.search.yahoo.com
	http://google.sina.com.tw
	http://google.pchome.com.tw
	http://www.openfind.com.tw
	http://search.yam.com
部落格	http://tw.blog.yahoo.com
	http://blog.sina.com.tw
	http://myblog.pchome.com.tw
家族	http://tw.club.yahoo.com
	http://club.yam.com
郵件系統	http://tw.f184.mail.yahoo.com
	http://mp.sina.com.tw
	http://mail.pchome.com.tw
	http://mail2000.com.tw
	http://freemail.yam.com

資料來源：本研究整理

針對以上所概述的相關系統，本研究亦參考 Tiwana 對知識相關平台分類為七層架構，將這些相關系統帶入這七層中，並將雲林縣政

府這些系統予以區分為七類，如圖 3。

圖 3 雲林縣政府相關系統對映至 Tiwana 的知識管理平台七層架構一覽表

介面層 瀏覽器
存取及授權層 安全，防火牆
合作、過濾及智慧層 搜尋引擎
應用層 電子表單，論壇，計畫管控，部落格，家族（會員）
傳輸層 公文交換，電子郵件
中介及既有系統整合層 精緻版 e 政府服務平台(<i>The Mini Governmental Service Platform, mGSP</i>)
儲存設備

資料來源：本研究整理

因為雲林縣政府進行知識管理系統的建置已進行多年，但一直沒有任何機會針對所建置的知識管理系統進行分析，本研究希望可以進一步瞭解雲林縣政府的知識管理系統，並據以制定建議政策。

乙、研究架構及研究步驟

在資料預先處理程序時，本研究先將冗餘資料去除，如 CGI 檔案、GIF 檔案、JPEG 檔案等去除，其次再將有遺漏值部份予以填補，最後將資料格式予以調整，如原本以日期顯示的欄位，因為需轉換成 Recency 與 Frequency，才能符合本研究資料探勘的標的。

在處理過程中，一般都會遇到不同使用者的問題，如在 DHCP 環境，同一部 PC 在不同時間上網，他的 IP 可能就不一樣，另外同一部 PC 可能會有不同使用者使用的問題。在本研究中，非常幸運可以減少上述兩項的影響，首先雲林縣政府都是採用固定 IP，沒有採用浮動 IP，其次雲林縣政府至 95 年為止，縣政府內的 PC 已達 1000 多部，幾乎達到每人一機，因此上述兩項對本研究影響不大。

因為研究的行為是使用者對知識管理系統的 Recency、Frequency，所以必須區分 Session，區分的用意在於區分這是否為使用者另外一個瀏覽行為，才能計算出 Recency 及 Frequency，而參考，以 30 分鐘為單位，區分不同的 session。

最後必須將來自不同系統的資料予以整合，雲林縣政府內部擁有一份對照表，指出 IP 與使用者的對照，因此可以靠著這份對照表，將 Proxy log 中的 IP 與人事資料結合起來，再針對這些員工背景資料予以分析。整理後的檔案成為表 4 所示。

表 4 經整理後的資料表

欄位名稱	資料屬性	說明
IDNO	CHAR	身分證號
IP	CHAR	電腦位址
SEX	CHAR	性別
BIRTHD	Integer	年齡
Age	CHAR	年齡級距
Marriage	CHAR	婚姻狀況
Unit	CHAR	單位
EDU	CHAR	教育程度
Title	CHAR	職稱
ADDR	CHAR	住址
TSERDY	Integer	服務年資
R	Integer	Recency
F	Integer	Frequency
M	Float	Monetary

而圖 3 的七層中，在本研究所能取得資料的範疇僅有「合作、篩選及智慧層」、「應用層」及「傳輸層」等三層，因此在接下來的討論中，均以此三層為主要討論目標。

在第一部份的分析，將採取不分層方式分析，而在第二部份的分析中，將區分「合作、篩選及智慧層」、「應用層」及「傳輸層」之資料區分為三類，並進行個別分析。

本研究進行的架構，茲以流程圖說明，如圖 4。由圖 4 所示，首

先需設計「分析層級程序法」之層級，這部份是利用現有文獻，以 Tiwana 對知識管理系統的分層方式，做為「準則」，其次以現有符合之系統做為評估的「替代方案」，設計好問卷即可以發予參與評估者，收回問卷時應特別注意其一致性， $CR \leq 0.1$ 才能採用。

爲了確定問卷調查受訪對象給予元素間的重要性具有一致性，「分析層級程序法」使用一致性比率 (Consistency Ratio, CR) 來測量其一致性，求 CR 之前需先求一致性指標 (Consistency Index, CI)，如(5)及(6)。

$$CI = \frac{\lambda_{\max} - n}{n - 1} \quad (5)$$

λ_{\max} 表示矩陣最大固有值， n 代表矩陣大小。

$$CR = \frac{CI}{RI} \quad (6)$$

RI 如表 5。

計算 CI 之前需先求得矩陣的最大固有值，而隨機性指標 (Random Index, RI) 的給予，係依 Satty 所給予的參考值，如表 5。計算結果如 $CR \leq 0.10$ 時則代表問卷調查結果是可以被接受的。

表 5 隨機性指標對照表

矩陣大小	1	2	3	4	5	6	7	8	9	10
RI 值	0.00	0.00	0.58	0.90	1.12	1.24	1.32	1.41	1.45	1.49

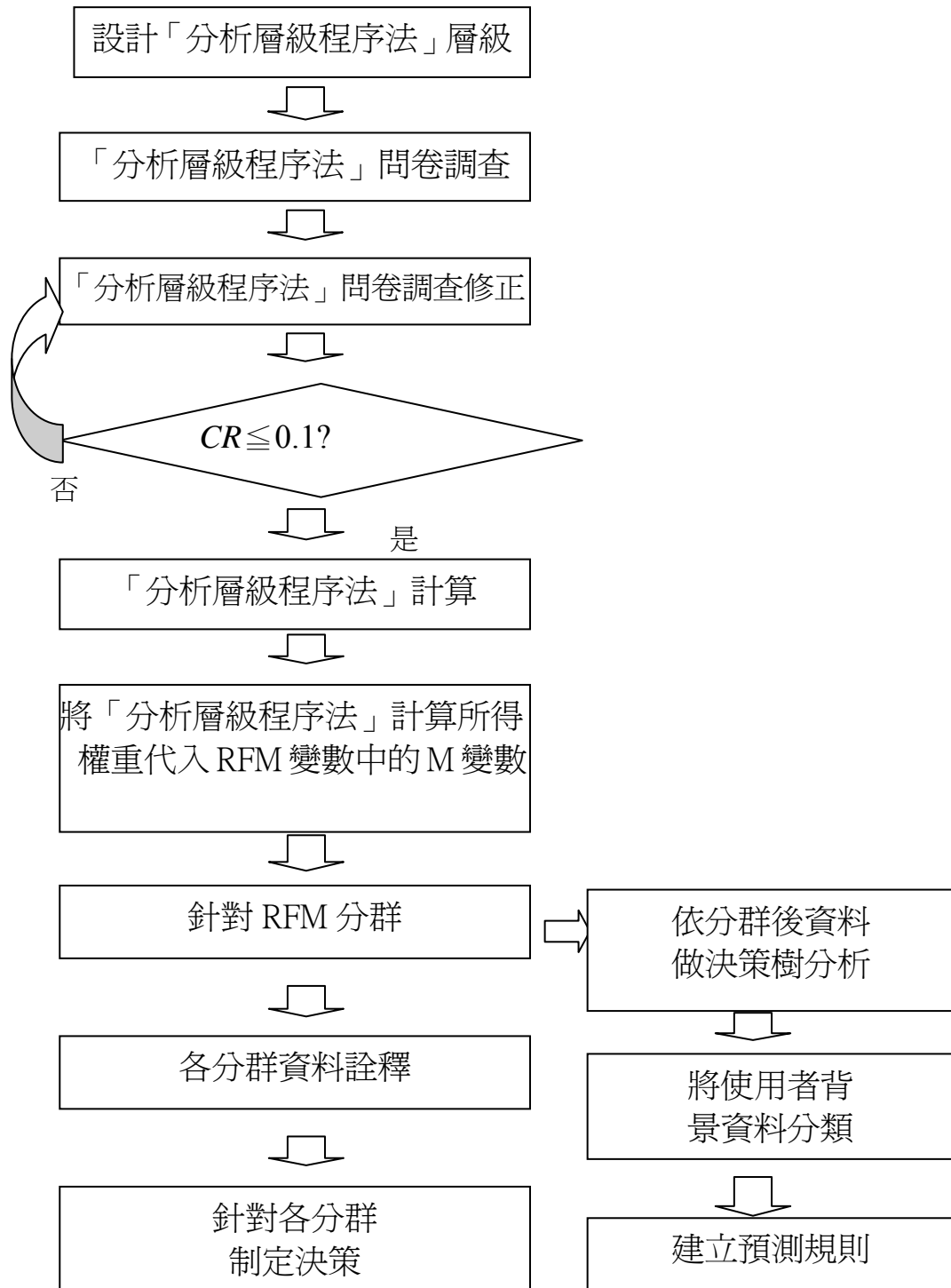
資料來源：多準則決策分析王小璠 (2005)，《多準則決策分析》

將「分析層級程序法」計算所得之權重代入 RFM 中的 M 變數中，再針對 RFM 三項變數以資料探勘中的分群方式區分群聚。針對這些群聚，本研究制定不同的應對決策，並分析使用者背景資料，預測使用者行爲，並分析該群聚使用者的背景資料，瞭解該群聚使用者的特性。

爲更清楚說明資料分析的流程，茲以圖 5 說明，首先以「分析層級程序法」及「顧客關係管理」的觀念，將「顧客生命週期價值」中

的 RFM 模式代入，但 M 變數的值改以「分析層級程序法」評估後的權重。要進行分群及分類之前，本研究先將使用者行為予以詮釋，再將資料以分群方法的 K-Means 演算法予以分群，找出那些是較具價值的員工，再將各群聚予以分類，進行背景預測及分類。

圖 4 研究流程圖



資料來源：本研究整理

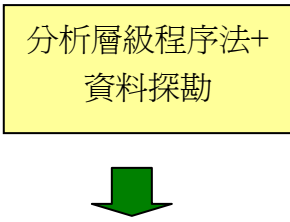
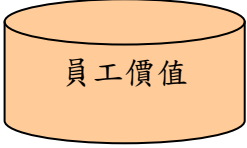
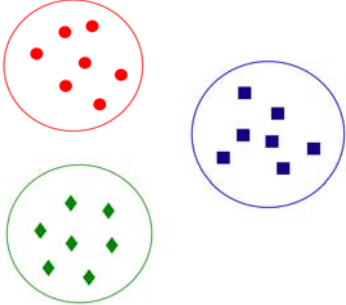
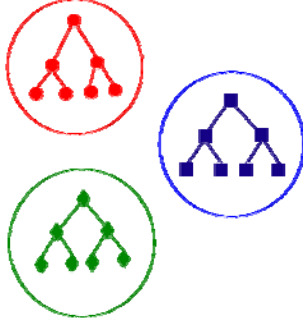
步驟	圖說	說明
1		<p>以「分析層級程序法」評估各系統權重，並以此權重代入 RFM 模式中的 M 變數</p>
2		<p>詮釋員工行為類別及其意義</p>
3		<p>分群方法 使用 K-Means 演算法，找出各群聚</p>
4		<p>分類方法 使用決策樹演算法預測員工背景並進行分類</p>

圖 5 資料分析流程圖示

資料來源：本研究整理

茲將資料探勘與資料整合情形再以下列步驟補充說明：

Step1:輸入資料。

輸入訓練樣本集合 T

$$T = \{\vec{x}_1, \vec{x}_2, \dots, \vec{x}_N\}, \quad x_i \in R^3, \quad \vec{x}_i = (x_{iR}, x_{iF}, x_{iM}), \quad i = 1, 2, \dots, N$$

Step2:進行 K-Means 分群演算法。

進行 KM 演算法，首先任選一資料為中心點 m_i

$$E = \sum_{i=1}^k \sum_{p \in C_i} |p - m_i|^2$$

樣本與 m_i 最近者，視為同一群，直至與一群集中心點不再有變動，演算法停止並輸出 K 個群。

$C = \{C_1, C_2, \dots, C_k\}$ ，且 $C_i \cap C_j = \phi$ ， $i \neq j$ ， $i, j = 1, 2, \dots, k$ 。

Step3 :計算樣本集合 T 的幾何中心 $(\bar{R}, \bar{F}, \bar{M})$ 。

計算 $\bar{R}, \bar{F}, \bar{M}$ 的公式分別如下

$$\bar{R} = \frac{1}{N} \sum_1^N x_{iR}$$

$$\bar{F} = \frac{1}{N} \sum_1^N x_{iF}$$

$$\bar{M} = \frac{1}{N} \sum_1^N x_{iM}$$

Step4 :計算各群幾何中心 $\bar{C}_{jR}, \bar{C}_{jF}, \bar{C}_{jM}$ 。

計算 C_j 之幾何中心各分量 $\bar{C}_{jR}, \bar{C}_{jF}, \bar{C}_{jM}$ 之公式如下：

令 $x_i = (x_{iR}, x_{iF}, x_{iM}) \in C_j$

$$(C_{jR}, C_{jF}, C_{jM}) = \frac{1}{\#(C_j)} \left(\sum_{x_i \in C_j} x_{iR}, \sum_{x_i \in C_j} x_{iF}, \sum_{x_i \in C_j} x_{iM} \right)$$

Step5 :計算各群屬於何種行為類型。

假設類型有八種 $\Omega = \{\Omega_1, \Omega_2, \dots, \Omega_8\}$ ，各類表示如下

$\Omega_1 = (H, H, H)$, $\Omega_2 = (L, H, H)$, $\Omega_3 = (H, H, L)$, $\Omega_4 = (L, H, L)$,

$\Omega_5 = (L, L, H)$, $\Omega_6 = (H, L, H)$, $\Omega_7 = (L, L, L)$, $\Omega_8 = (H, L, L)$

定義 C_j 的類別為 (C_j^R, C_j^F, C_j^M) 且 $C_j^R, C_j^F, C_j^M \in \{H, L\}$

計算 C_j 之公式如下

If $\bar{C}_{jR} \geq \bar{R}$ Then $C_j^R \leftarrow H$

If $\bar{C}_{jR} \leq \bar{R}$ Then $C_j^R \leftarrow L$

If $\bar{C}_{jF} \geq \bar{F}$ Then $C_j^F \leftarrow H$

If $\bar{C}_{jF} \leq \bar{F}$ Then $C_j^F \leftarrow L$

If $\bar{C}_{jM} \geq \bar{M}$ Then $C_j^M \leftarrow H$

If $\bar{C}_{jM} \leq \bar{M}$ Then $C_j^M \leftarrow L$

Step6:將各類型有相同之群聚，再加以整合，表示如下。

If $\{C_j^R, C_j^F, C_j^M\} \in \Omega_j$ and $\{C_{j+1}^R, C_{j+1}^F, C_{j+1}^M\} \in \Omega_j$
 then $\{S_j^R, S_j^F, S_j^M\} \leftarrow \{C_j^R \cup C_{j+1}^R, C_j^F \cup C_{j+1}^F, C_j^M \cup C_{j+1}^M\}$

Step7:將 *HHH* 與 *LHH* 類型整合為第一類型員工，將 *HLL* 與 *LLL* 類型整合為第四類型員工。

假設員工類型有四種 $\alpha_j = \{\alpha_1, \alpha_2, \alpha_3, \alpha_4\}$ ，而且
 $\alpha_1 = \{HHH, LHH\}, \alpha_2 = \{HHL, LHL\},$

$\alpha_3 = \{LLH, HLH\}, \alpha_4 = \{LLL, HLL\}$ ，整合方式表示如下

If $\{S_j^R, S_j^F, S_j^M\} \in \alpha_j$ and $\{S_{j+1}^R, S_{j+1}^F, S_{j+1}^M\} \in \alpha_j$
 then $\{\Gamma_j^R, \Gamma_j^F, \Gamma_j^M\} \leftarrow \{S_j^R \cup S_{j+1}^R, S_j^F \cup S_{j+1}^F, S_j^M \cup S_{j+1}^M\}$

Step8:將整合為第一類型員工及第四類型員工以決策樹演算法，建立預測規則。

假設 $\Gamma = \{S_1, S_2, \dots, S_j\}$

在 A 屬性中可以求得子集合 S，而我們可求得

$$E(A_k) = \sum_{j=1}^v \frac{S_{1j} + \dots + S_{mj}}{s} I(s_{1j}, \dots, s_{mj})$$

$$\text{其中 } I = (S_1, S_2, \dots, S_{mj}) = -\sum_{i=1}^m p_{ij} \log_2(p_{ij})$$

而我們可以求得 Gain(A) 最大值者為分支屬性，

$$\text{Gain}(A_k) = I(S_1, S_2, \dots, S_m) - E(A_k)$$

If the value of $\text{Gain}(A_k)$ is maxmun then output T_k

Step9:輸出決策樹結果。

Output a set of Decision tree

$$T = \{T_1, T_2, \dots, T_m\}$$

至於為何不在資料分群時，直接逕自設定兩群，為何需要一步一步分群，再予以整合?其原因在於分太多群，群聚之間發生太多相同類型，需要再做合併，以降低計算的複雜度。但也不宜分群太過簡略，雖然計算的複雜度降低，但不正確性將會提高。在本研究進行過程中，不斷測試結果，針對每種資料各有不同分群數量的設定，在本研究的不分群資料，如果分為 8 群或 12 群，合併後都只有 *LLL*、*HHH*、*HLL*、*LHH* 四種變化，但如果將資料再降低分至 6 群，將只能發現 *LLL*、*HLL*、*HHH* 三種變化，*LHH* 之行爲態樣資料將遺失，因此分群

之數量應測試至最適當情形。

而為提高預測準確度，且 *LLL* 與 *HLL* 本屬第四類型員工，*HHH* 與 *LHH* 屬於第一類型員工，可以再加以整合，除可提高正確性之外，在本研究中之預測模型，僅瞭解其屬於何種類型員工即可。

伍、研究結果

在設計「分析層級程序法」的層級時，本研究設計係以 Tiwana 所提出的知識管理架構的七層分類，將其中與本研究資料有關的三層拿來應用分析，這三層分別為「合作、篩選及智慧層」、「應用層」及「傳輸層」，將這三層視為「準則」，而「焦點」即設定為「促進地方政府知識管理」。

在設計完層級分析問卷後，本研究針對雲林縣政府內的五位員工進行調查，而這五位員工，其中有二位是參與知識管理系統開發時期之資訊單位人員、有二位是參與知識管理系統開發時期之受訪談人員或協助開發人員、有一位是單純知識管理系統之使用者。

經過「分析層級程序法」計算，將五位受訪者所評之權重以幾何平均法加總，得到表 11 結果。所評之權重，將代入 RFM 模式中的變數 M，以此變數 M 之數值做為各系統重要性之權重數，再將之以資料探勘方式分析其資料。在本研究中所使用的資料，代理伺服器的紀錄檔約有 180 萬筆資料，員工資料有 318 筆。

經過前述的資料預處理之後，以 RFM 為目標，進行資料分群，將經過賦權後的各系統使用紀錄予以分群為八個群聚，再將 R、F、M 三項變數各取其平均值，把八個群聚中的 RFM 各個的平均值減掉總平均值，據以區別該群聚中的 RFM 值是屬於「高於平均值」(H) 或「低於平均值」(L)，再針對各群的結果予以詮釋並制定政策，詳見表 6。傳統的 RFM 模式的操作方式不容易找出資料彼此間 RFM 接近之資料，因此近來在分析 RFM 模式時，資料分群的技术被引進，以提高其正確性。

表 6 資料詮釋及決策建議

類別	R	F	M	詮釋	決策建議
1	H	H	H	忠誠度高 價值性高	使用者經常上這些系統，使用系統之價值總值較高，為核心使用者。
	L	H	H		
2	H	H	L	忠誠度高 價值性低	經常上這些系統，使用系統之價值總值較低，應導引進階。
	L	H	L		
3	L	L	H	忠誠度低 價值性高	不常上這些系統，但使用系統之價值總值較高，應瞭解其真正需求，修正系統符合其使用。
	H	L	H		
4	L	L	L	忠誠度低 價值性低	不常上這些系統，使用系統之價值總值較低，應加強其基礎觀念，促進其使用，此類使用者應為初階使用者。
	H	L	L		

資料來源：本研究整理

不分層分析之資料在本研究中一律以 Case1 代稱之，而分層的資料分別是「合作及智慧層」類型的資料紀錄檔案則以 Case2 稱之、「應用層」類型的資料紀錄檔案則以 Case3 稱之、「傳輸層」類型的資料紀錄檔案則以 Case4 稱之。

在資料分群之前，先將上述不分層資料及分層資料先區分開，依序為 Case1 至 Case4，並且計算每一母體資料的 R、F、M 平均值，見表 7。這裡必須說明 R 變數及 M 變數，M 變數已知是由 AHP 評估後所得，而 R 變數是以整數取代日期，資料是以分鐘計算，數字的區間分別表示，見表 8。

資料分群的方式，會依資料量大小，事先預給六至八個不等之群聚，再將此群聚分別計算其 R、F、M 平均值，以「不分層 Case1」為例說明，表 7 中「不分層 Case1」的 R、F、M 平均值相較，可得知該群聚中的 R、F、M 平均值是屬於 High 或 Low，經整理後可得到表 9。但發現 C1、C4 及 C7 都屬於「HHH」型態，因此予以彙集整合分析，而 C6 及 C8 都是屬於「LHH」型態，故亦彙集一起，另外 C1 及 C5 都是屬於「LLL」型態，因此也是整合一起處理，至於 C3 為「HLL」型態，則 C3 單獨一個型態做為分析目標。

表 7 各類資料群其 RFM 平均值

各變數平均值 分層後紀錄	Recency	Frequency	Monetary
「不分層 Case1」	6731	33	2.496857
「合作及智慧層 Case2」	5819	5	0.107659
「應用層 Case3」	6577	35	2.988283
「傳輸層 Case4」	4797	3	0.243847

資料來源：本研究整理

表 8 Recency 日期時間表示法

數值區間	表示日期時間
1440 以下	表示 6 天以上
1441-2880	表示 5-6 天
2881-4320	表示 4-5 天
4321-5760	表示 3-4 天
5761-7200	表示 2-3 天
7201-8640	表示 1-2 天
8641 以上	表示 1 天以內

資料來源：本研究整理

表 9 「不分層 Case1」各群聚 RFM 劃分 High 或 Low 一覽表

變數 群聚	Recency	Frequency	Monetary
Cluster 1 (C1)	Low	Low	Low
Cluster 2 (C2)	High	High	High
Cluster 3 (C3)	High	Low	Low
Cluster 4 (C4)	High	High	High
Cluster 5 (C5)	Low	Low	Low
Cluster 6 (C6)	Low	High	High
Cluster 7 (C7)	High	High	High
Cluster 8 (C8)	Low	High	High

資料來源：本研究整理

本研究資料「不分層 Case1」資料計有 133 筆、「合作及智慧層 Case2」資料計有 118 筆、「應用層 Case3」資料計有 105 筆、「傳輸層 Case4」資料計有 23 筆。取樣方式以各群 70%隨機取出為訓練組資料，其餘 30%為測試組資料，分別做 5 次，每次均透過上述程序取出一組訓練組、一組測試組資料。

而「*HHH*」與「*LHH*」資料同屬第一類型員工(詳見表 6)，且「*LLL*」與「*HLL*」資料同屬第四類型員工(詳見表 6)，因此為使預測更精準，而且「*HHH* 與 *LHH*」或「*LLL* 與 *HLL*」沒有區分的必要，故只分第一類員工(Type1)及第四類(Type4)員工予以預測。

本研究從母體資料的各組資料中，如表 10，取出準確度最高那一組模型。故「不分層 Case1」因為其第三組資料模型的準確度為 64.58%，其準確度最高，故本研究採用之，其他如「合作及智慧層 Case2」的第三組資料模型之準確度為 67.58%、「應用層 Case3」的第二組資料模型之準確度為 76.79%、「傳輸層 Case4」模型的第三組資料模型之準確度為 69.62%，其準確度均為最高，故為本研究所採用。

表 10 各模型施以測試組資料後之準確度評估

母體資料 Case	測試組別	抽樣後測試正確率結果				
		第一組資料	第二組資料	第三組資料	第四組資料	第五組資料
1	訓練	69.9%	71.28%	68.04%	70.11%	71.59%
	測試	50%	41.03%	61.11%	58.7%	53.33%
	平均	59.95%	56.16%	64.58%	64.41%	62.46%
2	訓練	69.57%	70.45%	72.29%	71.26%	72.73%
	測試	57.69%	63.33%	62.86%	54.84%	56.1%
	平均	63.63%	66.89%	67.58%	63.05%	64.42%
3	訓練	75.34%	85%	73.33%	71.6%	77.78%
	測試	68.75%	68.57%	73.33%	79.17%	62.5%
	平均	72.05%	76.79%	73.33%	75.39%	70.14%
4	訓練	71.43%	87.5%	69.23%	75%	75%
	測試	66.67%	42.86%	70%	57.14%	57.14%
	平均	69.05%	65.18%	69.62%	66.07%	66.07%

在本研究中，將分析工作分成兩種程序進行，第一種是不分層的狀況下去分析，第二種是以分層的方式去分析。分層分析主要是再將所有資料區分為「合作及智慧層」、「應用層」、「傳輸層」等三類，各類再進行分群與分類。分層的用意在於將不同貢獻程度的系統分開討論；例如「合作層中相關系統的貢獻度」比「應用層中相關系統的貢獻度」低，自然不應混合一起分析其價值性，故第二種分析方式，將所有系統區分為「合作及智慧層」、「應用層」、「傳輸層」，再分別予以分群及分類等分析。

在資料分群之前，先將上述不分層資料及分層資料先區分開，依序為 Case1 至 Case4，並且計算每一母體資料的 R、F、M 平均值。

資料分群的方式，會依資料量大小，事先預給六至八個不等之群聚，再將此群聚分別計算其 R、F、M 平均值，以「不分層 Case1」為例說明，表 7 中「不分層 Case1」的 R、F、M 平均值相較，可得知該群聚中的 R、F、M 平均值是屬於 High 或 Low。

另外，因為「HHH」與「LHH」資料同屬第一類型員工(詳見表 6)，且「LLL」與「HLL」資料同屬第四類型員工(詳見表 6)，因此為使預測更精準，而且「HHH 與 LHH」或「LLL 與 HLL」沒有區分的必要，故只分第一類員工(Type1)及第四類員工(Type4)予以預測。

「不分層 Case1」的資料經過分群之後，可以整理得到「HHH」、「LHH」、「HLL」及「LLL」等四群資料，再依表 6 的定義，將上述四群予以歸納為第一類型員工及第四類型員工資料，最後再針對「不分層 Case1」群聚以決策樹演算法做分析，可以建立如圖 6 所列預測規則，可供今後新進員工進入縣府時，做為推動知識管理人員決策參考。

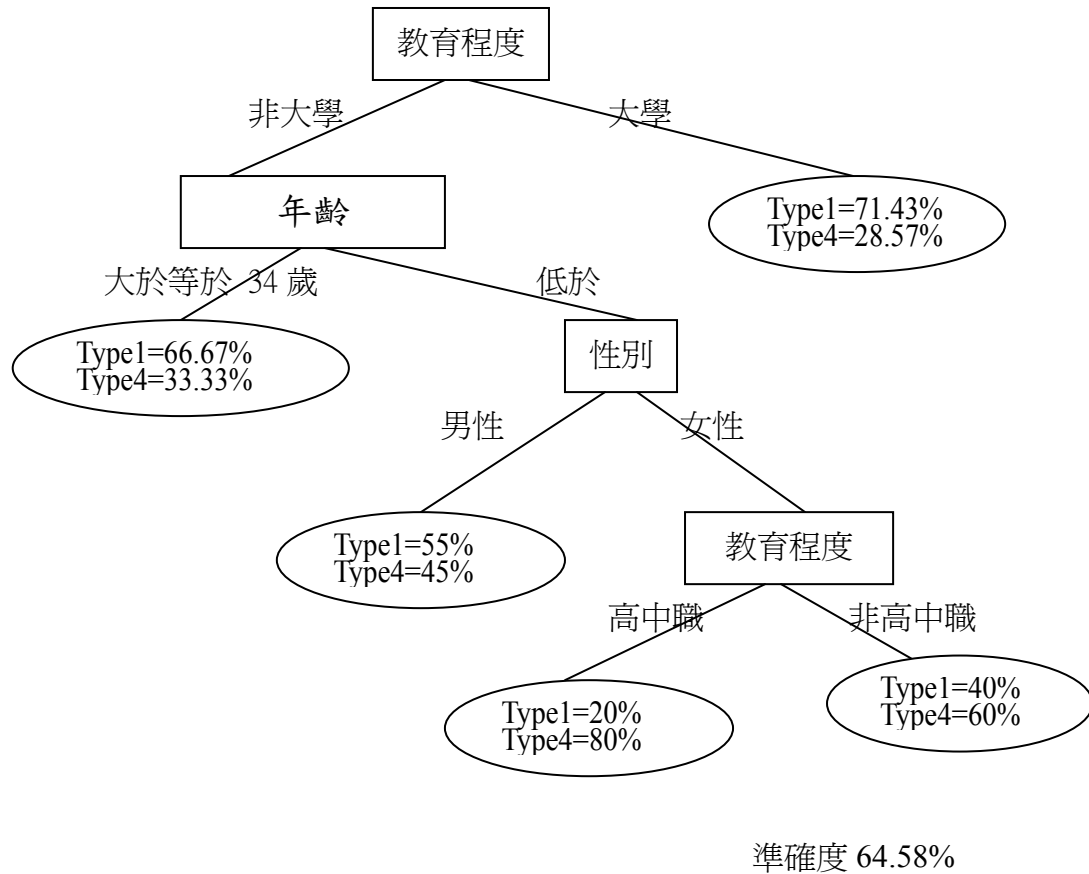


圖 6 「不分層 Case1」群聚之決策樹分析

但是在知識管理系統中，所有系統都有其不同功能及屬性，對組織的重要性也不同，第二種的分析是基於 Tiwana 的七層架構中的三層予以分別分析，因為在「分析層級程序法」分析中，以幾何加總後，可以得到表 11。可以發現其實經過評估後，三個「準則」還是有差別，權重最大的是應用層，這些包括有「電子表單」、「公告系統」、「管考資訊」、「網路公事包及文件管理」等等，其權重值大於 0.05，其次是傳輸層，最低的是合作層。因此針對不同重要性，做不同的瞭解是有其必要。

表 11 「替代方案」與「準則」權重

資訊能力	Alternative	Weight	Criteria	Weight
低	搜尋引擎	0.018114	合作層	0.11068
	通訊錄	0.007972		
	行事曆	0.009814		
	問卷調查	0.015068		

	待辦事項	0.028960		
	社群、留言、 員工交流園地	0.013447		
高	電子表單	0.088569	應用層	0.506385
	論壇	0.028179		
	公告系統	0.119083		
	管考資訊	0.071511		
	網路公事包 及文件管理	0.103169		
	部落格	0.026900		
	家族	0.024658		
中	郵件系統	0.070106	傳輸層	0.311562
	電子公文交換	0.189006		
	虛擬企業網路	0.039480		

首先在「合作、篩選及智慧層」分類中，一樣以分群方式可以找到三個群聚，分別是「*HHH*」、「*HLL*」及「*LLL*」，依照表 6 的定義，分別予以歸納為第一類型員工及第四類型員工等兩種，並以決策樹演算法建立預測規則，如圖 7。

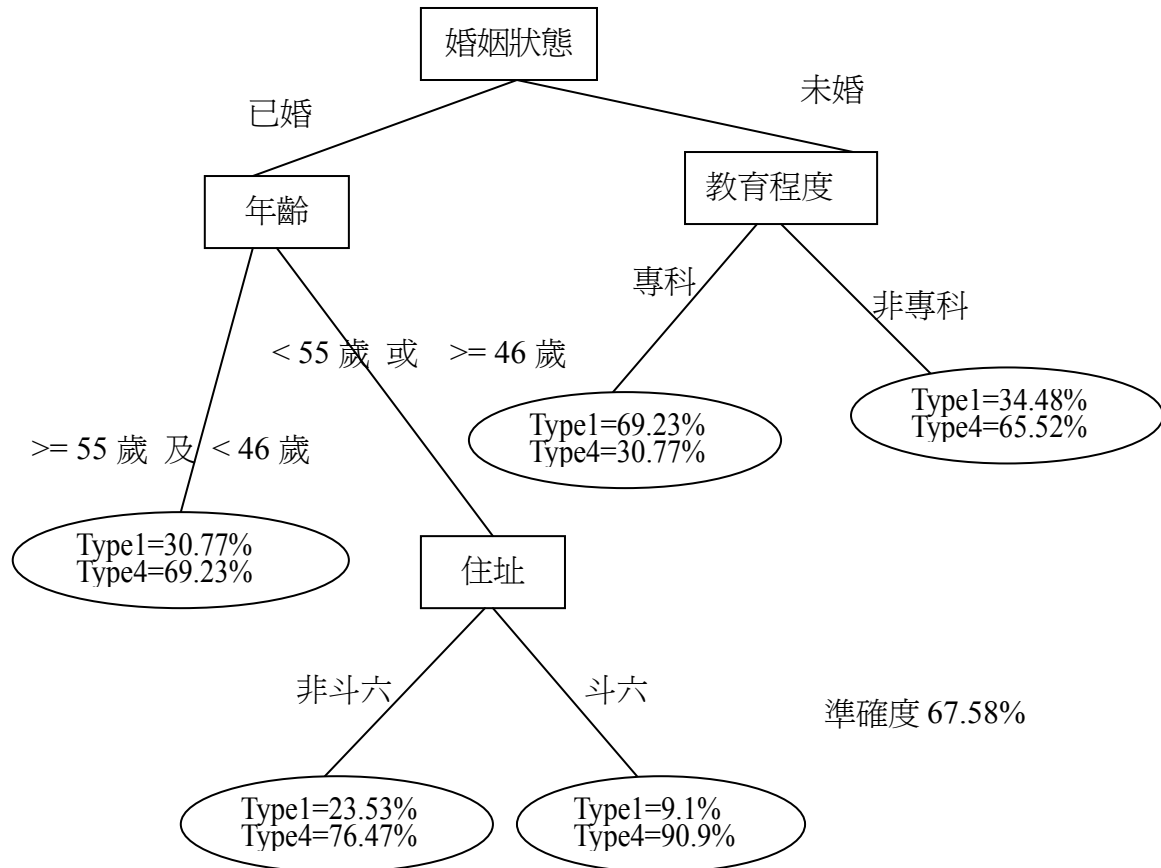


圖 7 「合作及智慧層 Case2」群聚之決策樹分析

在「應用層」的類別中，以資料探勘分群結果得到「*HHH*」、「*HLL*」、「*LHH*」、「*LLL*」等四群，依照表 6 之定義，予以歸納為第一類型員工及第四類型員工，並以決策樹演算法建立預測規則如圖 8。「應用層」的權重值是最高的，達 0.506385，可見在受訪者的考量下，「應用層」的相關系統如「電子表單」、「公告系統」、「管考資訊」、「網路公事包及文件管理」等等，對提升組織的知識管理非常重要，因此這類使用者對組織的重要性也非常高。

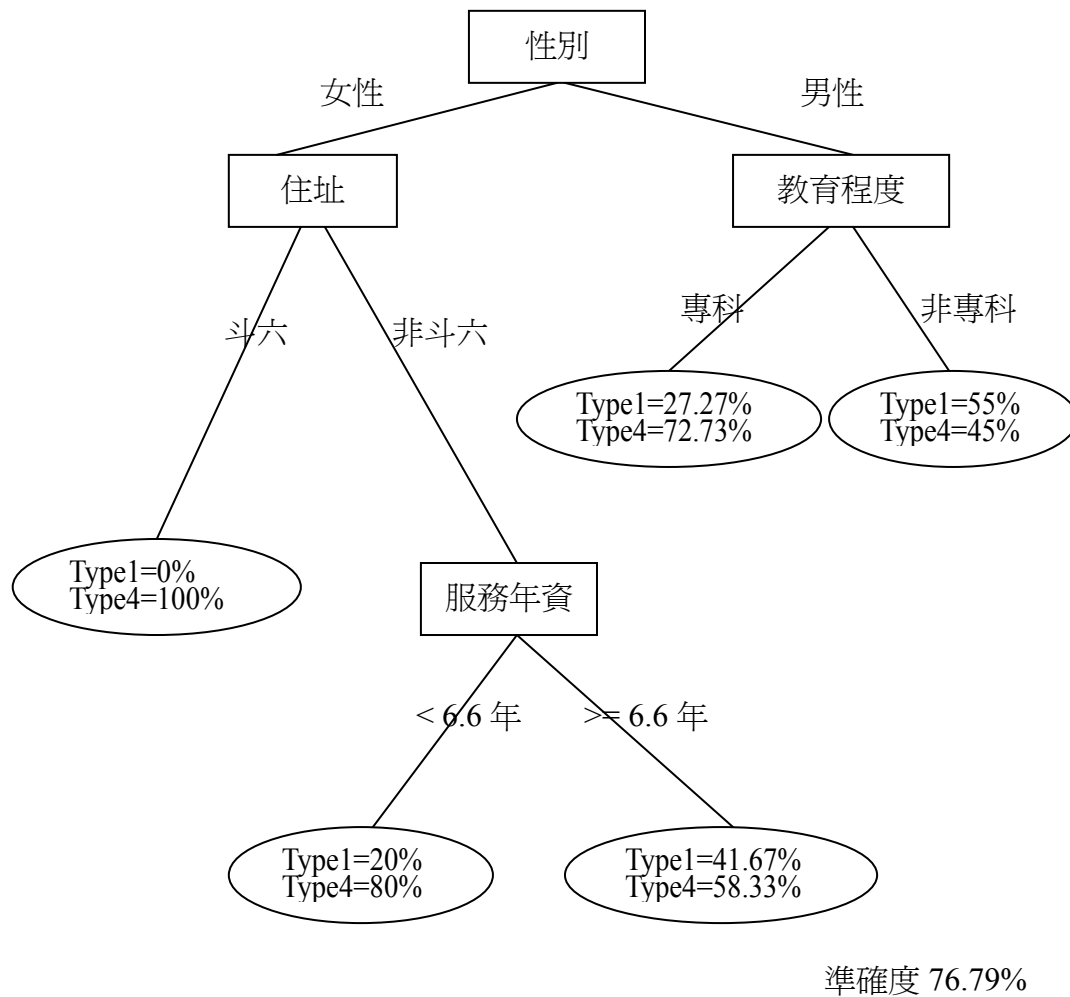


圖 8 「應用層 Case3」群聚之決策樹分析

受訪者對傳輸層的權重評分爲 0.311562，僅次於應用層，傳輸層中較重要的系統有「電子公文交換」及「郵件系統」。針對這類使用者予以分群，得到有「*HHH*」、「*HLL*」及「*LLL*」等三個群聚，歸納爲第一類型員工及第四類型員工之後，以決策樹演算法分析建立預測規則，如圖 9。至於圖 9 所示之傳輸層資料似乎過少，在準確度上應該尚有改進之空間。

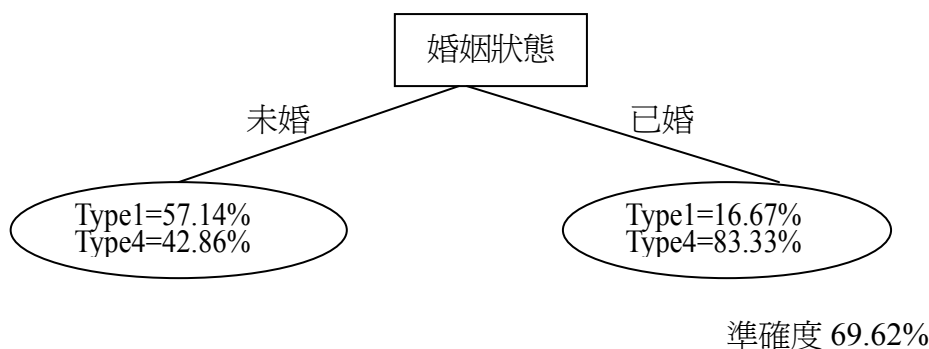


圖 9 「傳輸層 Case4」群聚之決策樹分析

陸、結論

在本研究取得資料時，雲林縣政府推動知識管理已多年，在本研究過程中，利用現有的代理伺服器紀錄資料及員工資料予以分析，做一次初步的檢視，讓推動知識管理人員或制定 IT 決策的人員可供參考，在本研究中所使用的資料，代理伺服器的紀錄檔約有 180 萬筆資料，員工資料有 318 筆，根據這些資料分析所得之結果，本研究提供雲林縣政府的建議，可分以下三個面向：

i. 組織策略方面

- ◆ 加強知識管理之基礎觀念，以凝聚所有員工之共識。
- ◆ 選擇副縣長或秘書長為單位的知識長(KEO)，並由人事處、主計處及計畫處資訊科，跨單位成立「知識管理環境推動小組」。
- ◆ 依照表 11 分析結果，建議針對具較高權重之系統，予以形塑強制性的需求文化。舉例以強制方式要求單位內所有不具機敏性資料，如公開之招標文件、規劃計劃案、統計報表等等，一律放置在「網路公事包及文件管理」；又如單位內充斥之紙本通告，可以強制一律以「公告系統」公布。

ii. 課程規劃方面

- ◆ 依照表 11 分析，因為「應用層 Case3」的相關系統對知識管理環境最具價值，因此其使用者亦為縣府推動知識管理過程中，較具價值之員工。可以參考圖 8 之預測規則，針對「應用層 Case3」類型之新進員工，建議予以菁英集訓，施以「網路進階蒐尋」、「協同合作相關系統應用」、「資料探勘」、「專案管理」等課程，讓這些高階使用者可以發揮其最大產出。單位內更應蒐集這類使用者名單，做為規劃設計團隊的儲備人才。
- ◆ 參考表 11 及圖 7，我們知道「合作及智慧層 Case2」的使用者所使用的系統，對知識管理環境比較不具價值性，因此建議參考圖 7 之使用者背景規則，針對以後此類之新進員工應該加強訓練，加速其應用知識管理工具技巧嫻熟度。

iii. 資訊管理系統發展策略方面

- ◆ 加強推動精緻版 e 政府服務平台(mGSP)，以 mGSP 整合相關系統，可以使更多知識管理系統整合在一起，更可以發揮相乘效果(Chang, 2007)。
- ◆ 參考表 11，可以發現「應用層 Case3」的相關系統，是組織內較具價值性之系統，縣府應該每年提撥一定額度，從自有預算持續發展這些系統，讓系統之使用價值加以發揮。
- ◆ 同樣地，參考表 11，建議檢討「合作及智慧層 Case2」相關子系統，針對不具重要性且使用率較低者，應予以汰換，除可去蕪存菁之外，亦可節省縣府不必要的系統維護費用。例如通訊錄及行事曆子系統，權重小於 0.01，比其他系統之權重還低，而且使用率不高，建議可以去除。

參考文獻

- 王小璠 (2005), 《多準則決策分析》, 臺中: 滄海書局。
- 戴文坡、普賽克 (2001), 《知識管理: 企業組織如何有效運用知識》, 臺北: 中國生產力中心。
- Agalgaonkar, A. P., S. V. Kulkarni and S. A. Khaparde (2006), “Evaluation of configuration plans for DGs in developing countries using advanced planning techniques,” *IEEE Transaction on Power Systems*, 21, 973-981.
- Buckinx, W. & D. V. d. Poel (2005), “Customer base analysis: partial defection of behaviorally loyal clients in a non-contractual FMCG retail setting,” *European Journal of Operational Research*, 164, 252-268.
- Bult R., & T. J. Wansbeek (1995), “Optimal selection for direct mail,” *Marketing Science*, 14, 378-394.
- Chang, T.-K., S.-S. Lin and H.-H. Tsai (2007), “The mGSP applications for local government in Taiwan – a case study for YUNLIN county government,” in *Proceedings of the Second Taiwan Conference on Software Engineering*, Taipei, 220-224.
- Dey, P. K.(2004), “Decision support system for inspection and maintenance: a case study of oil pipelines,” *IEEE Transactions on Engineering Management*, 51, 47-56.
- Dunham, M. H. (2002), *Data Mining Introductory and Advanced Topics*, Prentice Hall.
- Greiner, M. A., J. W. Fowler, W. M. Carlyle and R. T. McNutt (2003), “A hybrid approach using the Analytic Hierarchy Process and integer programming to screen weapon systems projects,” *IEEE Transactions on Engineering Management*, 50, 192-203.
- Huiying, Z. & L. Wei (2004), “An intelligent algorithm of data pre-processing in web usage mining,” in *Proceedings of the 5th World Congress on Intelligent Control and Automation*, Hangzhou, 3119-3123.
- King, S. F. (2007), “Citizens as customers: exploring the future of CRM in UK local government,” *Government Information Quarterly*, 24, 47-63.
- Liu, D.-R. & Y.-Y. Shih (2005), “Integrating AHP and data mining for product recommendation based on customer lifetime value,”

- Information & Management*, 42, 387-400.
- Marcus C. (1998), "A practical yet meaningful approach customer segmentation," *Journal of Consumer Marketing*, 15, 494-504.
- Rust, R. T., & P. C. Verhoef (2005), "Optimizing the marketing interventions mix in intermediate-term CRM," *Marketing Science*, 24, 477-489.
- Shih, Y.-Y. & C.-Y. Liu (2003), "A method for customer lifetime value ranking – combining the analytic hierarchy process and clustering analysis," *Database Marketing & Customer Strategy Management*, 11, 159-172.
- Sloane, E. B. (2004), "Using a decision support system tool for healthcare technology assessments," *IEEE Engineering in Medicine and Biology Magazine*, May/June, 42-55.
- Srivastava, J., R. Cooley, M. Deshpande and P.-N. Tan (2000), "Web usage mining: discovery and applications of usage patterns from web data," *SIGKDD Explorations*, 1, 1-12.
- Themistocleous, M., Z. Irani and P. E. D. Love (2005), "Developing E-Government integrated infrastructures: a case study," in *Proceedings of the 38th Hawaii International Conference on System Science*.
- Tiwana, A. (2002), *The Knowledge Management Toolkit*, New Jersey: Prentice Hall PTR.
- Wang, K., C. K. Wang and C. Hu (2005), "Analytic Hierarchy Process with fuzzy scoring in evaluating multidisciplinary R&D projects in China," *IEEE Transactions on Engineering Management*, 52, 119-129.
- Xu, M., Y. Qiu and J. Qiu (2003), "Mining for profitable customers," in *Proceedings of the International Conference on Information Technology: Computers and Communications*.
- Yuan, F., L.-J. Wang and G. Yu (2003), "Study on data preprocessing algorithm in WEB LOG mining," in *Proceedings of the Second International Conference on Machine Learning and Cybernetics*, Xi'an, 28-32.

The Assessment of Information System Based on Knowledge Management : A Case Study for Yunlin County Government

Ting-Kuei Chang

Dept. Information Management
Yunlin Univ. of Sci. & Tech.

Hung-Hsu Tsai

Dept. Information Management
National Formosa University

ABSTRACT

Since 2003, The Yunlin County Government starts to use a knowledge management system for document retrieval, electronic form and announcement systems etc. Most of the information systems can provide the information of user behavior by analyzing user's logs. However, while Yunlin County Government promotes the knowledge management system, the related analytical research of the date of users' background and users' logs are yet still lacking. Therefore, the research employs data mining techniques to analyze proxy server records and employee's data, and thereby to excavate the information of employee's behavior of using the system. As a result, managers of the Government can refer the results to make appropriate information system management policies for the employee in the Government.

This research proposes a novel technique which simultaneously employs the Analytic Hierarchy Process and the Data Mining in order to specify the staff who frequently use the system in Yunlin County Government. Furthermore, we take advantage of the Analytic Hierarchy Process to find out the weighting of Information System, and combined the Data Mining with the purpose of revealing the user's behavior. Finally, the weighting exposes the significance of Information System, also, the mining results provide the knowledge of system's user behavior for manager to make strategies for Information System.

KEYWORDS: Analytic Hierarchy Process, Customer Lifetime Value, Data Mining, Knowledge Management, YUNLIN County Government