

資料探勘於預測汽車維修項目之研究

楊珮珊

崑山科技大學資訊管理所

研究生

e-mail :

yukimei888@gmail.com

王平

崑山科技大學資訊管理系主任

兼所長

e-mail :

pingwang@mail.ksu.edu.tw

摘要

汽車保修之資料分析與汽車維修市場的產業成長息息相關。先前研究已運用 R 語言初步統計分析汽車保修項目與維修金額之關聯；本研究依據先前研究結果，針對汽車故障問題的損壞零件做進一步分析，使用 R 語言與 Weka 資料探勘工具進行交叉比對，目標透過資料探勘 (data mining) 中的迴歸分析 (regression analysis) 與決策樹 (decision tree)，將主要損壞零件間進行資料關聯分析，最後探討損壞零件間的關聯規則，提供維修廠做檢測參考。

關鍵詞：資料探勘、汽車維修、R 語言、決策樹、迴歸分析

Abstract

Data analysis is crucial for vehicle maintenance industry as the increasing of growth of vehicle maintenance market. Previous study used the data mining tool with R package to perform an association analysis regarding the relationship between vehicle maintenance items and expense from statistical data. Based on the research results of previous study, this study conducts experiments on correlation analysis to examine failure problem with the relevant vehicle components in further by using R package and Weka for cross-comparison purpose. Furthermore, we use the regression analysis and decision tree in data mining approaches to analyze the relationship among main vehicle damaged components and discover the association rules for damage reasons to provide a reference for future vehicle maintenance.

Keywords: Data mining、Vehicle maintenance、R、Decision tree、Regression analysis

1. 研究動機及目的

在生活中，汽車是相當重要的交通工具，但是大部分車主無法預知零件壽命，當交通途中發生零件損壞，即造成無法正常使用的困擾。現今市場上，原廠有保固期間的客源，而私人維修廠卻有相對於原廠較低價格能夠吸引顧客，並且競爭激烈。

現今資料探勘在各種領域使用廣泛，本研究使用資料探勘技術於汽車維修業進行預測分析，提供決策者進行精確、快速且可靠的維修服務，藉此節省時間與金錢成本，提升汽車維修的效率與車主的滿意度，讓維修廠的經營達到永續穩定成長。

由於本研究數據來自單一維修廠，故會受到地區性限制而產生不客觀之分析結果。

本研究結構如下：第二節將概述汽車維修產業與資料探勘技術及相關應用。第三節說明研究架構與資料前處理以及所使用的相關工具。第四節則說明研究結果並探討。最後針對本研究做結論與未來研究方向討論。

2. 相關文獻

傳統資料分析的技術難以從大量資料中萃取出有意義的規則(rule)，因此資料分析的技術和應用在過去幾年內快速發展。

而汽車保有量的增減也與汽車維修市場的產業息息相關，一般新車使用 6~7 年後，車上零組件開始老化、效能開始衰退，車輛故障率相對慢慢增加，可見汽車維修保養的重要性。

2.1 汽車維修業概述

國內汽車維修產業已經非常成熟且技術密集，台北市汽車保養商業同業公會[2]提出汽車

維修業之名詞定義。

其中清潔定義是指清潔污物之作業；潤滑：指更換或補充潤滑油脂；檢查：以目視判斷或儀器測試；調整：就汽車各部位機件以工具或儀器調諧使符合規範；維護：對汽車各部位機件局部分解並清潔、潤滑、檢查、調整及更換消耗性零件；總成更換：將汽車零、組件整體更換，使恢復正常之作業；校正：將不符合規範或變形部位，以各種設備修理使符合規範；翻修：將各單位總成細部分解，徹底修理或更換，含板金及塗裝。

2.2 資料探勘技術與應用

資料探勘（data mining）是指從大量資料中自動搜尋隱藏於其中特殊關聯性的資訊過程，由許多不同的軟體和技術組合而成，通過統計（statistics）、線上分析處理（on-line analytical processing）、情報檢索（information retrieval）、機器學習和模式識別（pattern recognition）等諸多方法來實現目標，並非運用單一技術就能解決所有問題。

相關技術區分為以下幾種類別：分類（classification）、推估（estimation）、預測（prediction）、關聯分組（affinity grouping）、群集（clustering），各功能簡要說明如表 1 所示。

表 1：資料探勘功能

功能	說明	應用方法
分類 (classification)	按照分析對象的屬性分類加以定義，建立類組(class)	決策樹、貝式網路
推估 (estimation)	根據連續數值之相關屬性資料，以獲某一屬性未知之值	相關分析、迴歸分析及類神經網路
預測 (prediction)	根據對象屬性之過去觀察值來推估該屬性未來之值	迴歸分析、時間序列及類神經網路

關聯分組 (affinity grouping)	透過資料分析在同一時間發生的事件，並呈現搜尋結果的規則	關聯分析
群集 (clustering)	將異質母體中區隔為較具同質性之群組 (clusters)	K-means、agglomeration

資料來源：本研究整理

資料探勘可應用於各行業，如電信業可進行客戶流失分析、銷售預測和詐欺偵測；零售業可進行客戶忠誠度、購物籃分析、交叉行銷、定價分析和銷售預測；而製造業可進行客戶貢獻度分析、行銷績效分析、品質管制、存貨分析和生產分析。除了各行業，尚有許多方面可應用資料探勘技術，如同我們生活息息相關的醫療保健、生物科技、航太空業、環境等資料皆可做分析，由此可見資料探勘的廣泛。

2.3 資料探勘架構

首先對研究目標與問題領域有清楚的定義與架構。龐大的資料須先經過整理以利後續分析，再針對問題本質使用合適工具進行資料分析，以發現有用的資訊。根據所得到的資訊擬定適當的行動，最後評估探勘的成效。資料探勘架構如圖 1 所示。

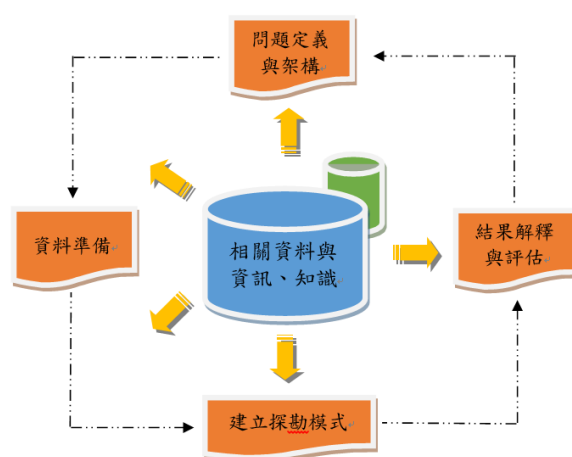


圖 1：資料探勘架構

資料來源：本研究重繪

3. 研究架構與方法

目前常見的分析工具相當多，本研究使用 R 語言與 WEKA 兩個工具做比對，接下來將做簡單介紹。

3.1 R 語言

R 語言是專為機率統計而設計的開放原始碼軟體。R 語言具有完整程式語言功能，能夠直接對應統計領域的變數型態，涵蓋數百個擴充套件(packages)供使用者安裝，可撰寫的程式語言包含 Python、Java、C、Fortran 等，可在 Windows、Mac、Unix、Linux 系統執行。

R 語言繪圖功能非常強大，運用資料視覺化(data visualization)能夠讓人更容易且有效地理解，在資料分析中，掌握資料視覺化的技巧是一項很重要的能力。

3.2 Weka

Weka(Waikato Environment for Knowledge Analysis)是由紐西蘭懷卡托(Waikato)大學所開發。以 Java 為基礎的資料探勘與 Machine learning 軟體，可在 Windows、Linux 與 Macintosh 系統執行，使用者下載安裝後即可使用。

Weka 整合大量的資料探勘演算法，提供 API 讓程式設計者使用，且具有相當完整的功能供使用者做分析，包括：資料前處理(Preprocess)、分類分析(Classify)、分群分析(Cluster)、關聯規則(Associate)，還有豐富的視覺化輸出。

3.3 研究資料

本研究以南部一間個人經營汽車維修廠為對象，使用所提供之顧客進廠維修資料做資料探勘分析，期間為民國 100 年 1 月初至民國 105 年 1 月底，共 5 年期間的維修紀錄為實證研究對象，資料包含客戶名稱、車牌、廠牌車型、車齡、里程數、C.C 數、維修品項、零件單價、維修日期等紀錄進行數據分析。

3.4 研究方法

依據問題尋找影響維修項目較顯著的因子。使用資料探勘中的迴歸分析建立預測模型，運用決策樹將主要損壞零件進行資料分析，目標屬性設定維修零件，廠牌、里程數、

車齡為預測屬性，有效率預測出維修項目並分類葉節點。圖 2 為本研究資料分析流程。



圖 2：資料分析流程

資料來源：本研究整理

3.5 資料前處理

維修廠原始提供客戶資料 434 筆、廠商資料 408 筆、零件資料 8878 筆、車輛入廠記錄 3307 筆、車輛維修內容 17914 筆、客戶付款資料 3071 筆，總共 34012 筆資料。

經資料整理後，僅使用客戶資料和主要維修零件資料共 9286 筆，由於客戶資料確實輸入的筆數較少，將兩個資料表整合後剩餘 2531 筆，再整理出資料異常值、離群值，並理解資料結構與摘要後進行數據分析，分析過程中反覆重新整理資料，最後可分析資料筆數僅 1463 筆。

資料探勘的表格屬性包含維修零件名稱、維修金額、車齡、里程數、廠牌，整理如表 2。

表 2 探勘資料

資料欄位	欄位解釋
ItemName	維修零件名稱
SaleTotal	維修金額
CarYear	車齡
LastMove	里程數
CarFactor	廠牌

資料來源：本研究整理

整理出的廠牌共有 13 種，分別為 A 廠牌、B 廠牌、C 廠牌、D 廠牌、E 廠牌、F 廠牌、G 廠牌、H 廠牌、I 廠牌、J 廠牌、K 廠牌、M 廠牌和 Others 其他廠牌，如表 3 所示。

表 3 廠牌資料

A	B	C	D	E
F	G	H	I	J
K	M	Others		

資料來源：本研究整理

由於維修項目種類過多，本研究最後整理出 20 項主要維修項目，刪除維修項目較少且不影響行車的資料，原 1463 筆資料刪除後剩餘 601 筆，並將維修零件資料重新命名以利分析，維修零件名稱對照如表 4 所示。

表 4 維修零件名稱

代號	欄位解釋	代號	欄位解釋
A1	下支臂	B1	傳動軸
A2	時規皮帶	B2	煞車分泵
A3	引擎腳	B3	機油濾清器
A4	水泵浦	B4	避震器
A5	水箱	B5	離合器
A6	火星塞	B6	煞車來令片
A7	外部配件皮帶	B7	汽車保固
A8	時規惰輪	B8	油封
A9	發電機	B9	空氣濾清器
A10	軸承	B10	繼電器

資料來源：本研究整理

4. 結果與討論

本研究首先使用 R 語言做資料型態分析，繪製出維修零件的直方圖與廠牌數的圓餅圖，之後分析車齡、里程數與維修金額間的關係，整理出各廠牌平均維修金額並繪製盒鬚圖，以及里程數的分佈圖，最後使用 R 語言與 Weka，做迴歸分析 (regression analysis) 與決策樹 (decision tree) 進行交叉比對。

4.1 研究結果整理

使用 R 語言裡的 summary() 函數查看資料型態，得到樣本數據的最小值、最大值、四分

位數以及平均值，如圖 3 所示。

```
> summary(test1)
ITEMName      SalTotal      CARYEAR      LASTMOVE      CARFACTOR
B9      :105   Min.    : 0      Min.    : 2.00   Min.    : 15750   M      :147
A7      : 67   1st Qu.: 300    1st Qu.:10.00  1st Qu.: 83386   B      : 79
B3      : 52   Median: 800    Median:11.00  Median:141002   OTHERS : 67
B6      : 49   Mean    :1337    Mean    :12.89  Mean    :147321   J      : 65
B7      : 40   3rd Qu.:1550    3rd Qu.:16.00  3rd Qu.:211334   C      : 49
A6      : 34   Max.    :10000    Max.    :28.00  Max.    :384221   F      : 45
(other) :253                                     (other) :148
```

圖 3：資料型態

資料來源：本研究整理

可得知維修零件項目前幾名最多為 B9 (空氣濾清器)105 筆，最少筆數沒有顯示，另使用 ggplot2() 函數繪製出所有維修零件直方圖如圖 4 所示。

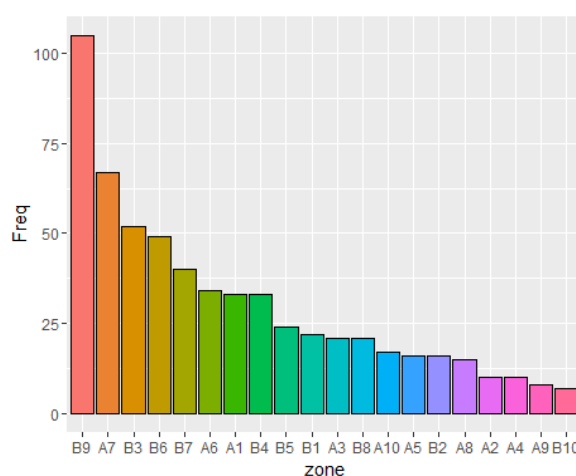


圖 4：維修零件數

資料來源：本研究整理

零件維修金額平均於 1300 元；進廠維修車齡平均為 13 年，最少為 2 年，最多為 28 年；維修車輛平均里程數為 14.7 萬，最少為 1.5 萬，最多為 38.4 萬；維修車輛前幾名最多為 M 廠牌 147 筆，最少筆數沒有顯示，另使用 pie() 函數繪製出所有廠牌圓餅圖如圖 5。

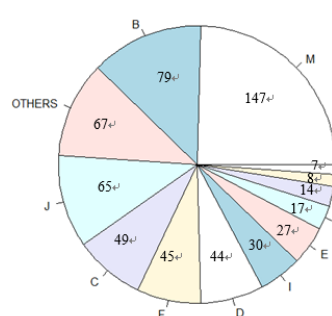


圖 5：廠牌數

資料來源：本研究整理

預分析車齡、里程數與維修金額間的關係，使用 `lattice()` 函數，車齡設定為 Z 軸，維修金額和里程數設定為 X 和 Y 軸，得知車齡與里程數間較有相關性，如圖 6 所示。

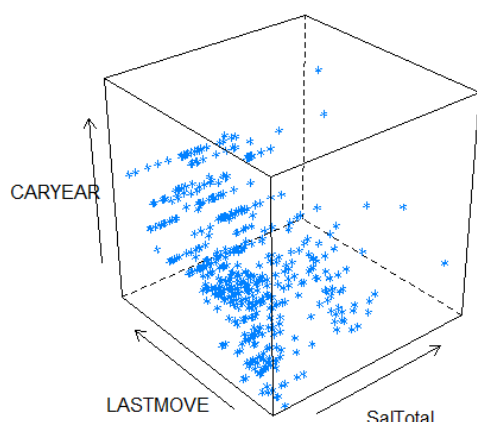


圖 6：車齡&里程數&維修金額關係

資料來源：本研究整理

應用 `summarize()` 函數，依據資料行分析出各廠牌汽車平均維修金額如圖 7，可發現 A 廠牌零件平均維修金額 2107 元較高，而 C 廠牌平均維修金額 895 元較低。

```
> summarize(group_by(te:
# A tibble: 13 x 2
  CARFACTOR  平均
  <fctr>      <dbl>
1 A      2107.857
2 B      1254.557
3 C       895.102
4 D      1248.182
5 E      1687.037
6 F      1760.000
7 G      1640.000
8 H      1232.353
9 I      1841.000
10 J      1790.769
11 K      1600.000
12 M      1102.568
13 OTHERS 1045.075
```

圖 7：各廠牌平均維修金額

資料來源：本研究整理

再應用 `boxplot()` 函數取 `log10` 以盒鬚圖呈現，圖 8 可看出 A 廠牌平均維修金額較高，而 C 廠牌平均維修金額較低，G 廠牌與 K 廠牌維修金額於圖 7 是相對較高的，但由於整體分析數量較少，因此繪製出的圖較為異常。

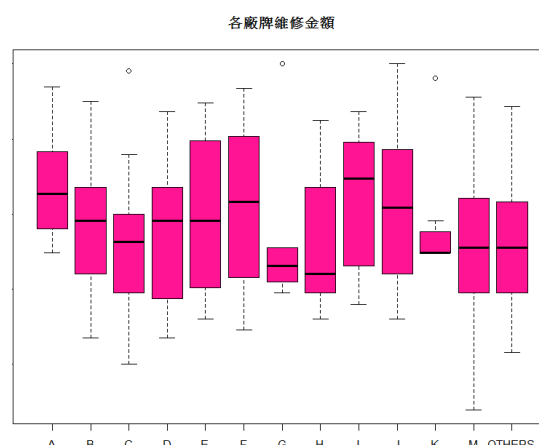


圖 8：各廠牌平均維修金額盒鬚圖

資料來源：本研究整理

應用 `hist()` 函數，設定資料中的里程數繪製出直方圖，數據以公里（萬）顯示，圖 9 可發現進廠維修汽車里程數大多分佈於 20 幾萬公里。

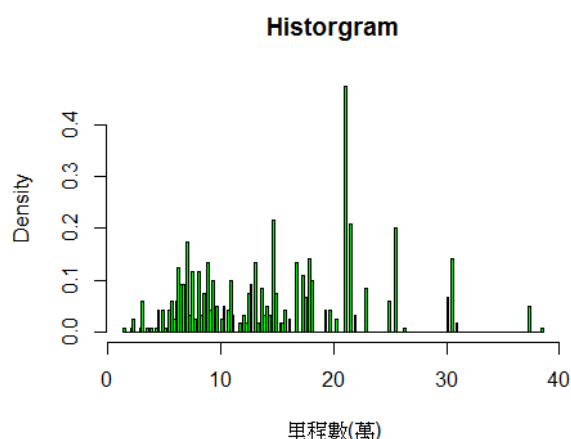


圖 9：里程數分佈圖

資料來源：本研究整理

4.2 R 語言迴歸分析

對樣本資料而言，皮爾森積矩相關係數 (Pearson Product Moment correlation coefficient) 的定義如下：

γ_{XY} = 樣本相關係數， $S_X = X$ 的樣本標準差， $S_Y = Y$ 的樣本標準差。

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

試分析里程數與車齡間有無關聯性，應用迴歸分析建立預測模型，將車齡設定為 x ，里程數設定為 y ，運用 `plot` 函數取 `log` 繪製散佈圖，並將線性迴歸設定為新模型後，輸入 `summary()` 函數即顯示模型相關資訊，如圖 10 所示，圖中可以看出 X (因子) 變量的顯著性，* 數越高代表越顯著。

```
> summary(fit)

Call:
lm(formula = log10(y) ~ log10(x))

Residuals:
    Min       1Q   Median       3Q      Max
-0.63071 -0.14303  0.03867  0.17308  0.42490

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.46740    0.04728   94.50  <2e-16 ***
log10(x)      0.59892    0.04350   13.77  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2108 on 598 degrees of freedom
Multiple R-squared:  0.2407, Adjusted R-squared:  0.2395
F-statistic: 189.6 on 1 and 598 DF, p-value: < 2.2e-16
```

圖 10：R 語言車齡與里程數迴歸分析

資料來源：本研究整理

R-squared 值代表迴歸的關聯度，值越高模型解釋能力越好，圖 11 可以清楚看到此模型 R-squared 值為 0.2407，表示車齡可以解釋里程數 24 % 的變異量，雖然表面上解釋能力不高，但是 X (因子) 變量已經顯著了，所以迴歸模式之有效性已達成，因此車齡與里程數存在著關聯，大部分車齡越高里程數也越高。

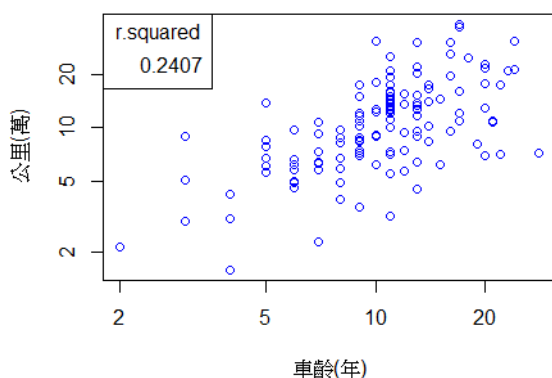


圖 11：R 語言車齡與里程數迴歸分析結果

資料來源：本研究整理

4.3 R 語言決策樹

本研究使用 R 語言裡的 `rpart()` 函數分類迴歸樹。

分類迴歸樹 (CART, Classification and Regression Tree) 是以反覆運算的方式，由根部開始反覆建立二元分支樹，直到樹節點中的同質性達到某個標準，或觸發反覆運算終止條件為止。CART 的應變數欄位可以是類別型資料或數值型資料。

以下用樣本資料 `test1.csv` 做分析，共有 5 個欄位，包括 `ItemName` (維修零件名稱)、`SaleTotal` (維修金額)、`CarYear` (車齡)、`LastMove` (里程數)、`CarFactor` (廠牌)。以下簡單介紹所輸入的程式：

- ◆ `n=0.3*nrow(test1)`
解釋：nrow(test1) 傳回 test1 的資料數
- ◆ `test.index=sample(1:nrow(test1),n)`
解釋：30% 測試資料(600 筆資料總數 * 0.3 = 180 筆)
- ◆ `test1.train=test1[-test.index,]`
解釋：訓練資料
- ◆ `test1.test=test1[test.index,]`
解釋：測試資料
- ◆ `test1.tree=rpart(ITEMName~SalTotal+CARFACTOR+LASTMOVE+CARYEAR, data= test1.train)`
解釋：建立 test1.tree 訓練資料

接下來計算訓練樣本與測試樣本的預測正確率後，得到結果：訓練樣本預測正確率為 81%，而測試樣本預測正確率為 72%，如圖 12 所示。

```
> sum(diag(cmatrix))/sum(cmatrix) #預測正確率
[1] 0.815
> sum(diag(cmatrix))/sum(cmatrix) #預測正確率
[1] 0.7349398
```

圖 12：R 語言決策樹預測正確率

資料來源：本研究整理

最後使用 `rpart()` 函數繪製出決策樹如圖 13，依據圖 13 決策樹路徑上的節點與分支，將這些規則製成表格如表 5 與表 6。

可以分成維修金額大於 465 元與小於 465 元兩種類型，表格中的 X 代表沒有出現的分支，若有一欄位整排都是 X ，代表沒有出現該節點。

表 5 分類結果(維修金額> 465 元)

ItemName (維修零件名稱)	SaleTotal (維修金額)	CarYear (車齡)	LastMove (里程數)	CarFactor (廠牌)
B4	高於 3850 元	X	高於 16.8 萬	X
B7	高於 3850 元	X	低於 16.8 萬	X
A3	1875~3850 元	X	X	D、I
B4	1875~3850 元	X	X	不是 D、I
B4	低於 1875 元	X	X	不是 B、C、D、E、F、 H、I、J、M、Others
B5	低於 1875 元	X	X	不是 B、C、D、E、F、 H、I、J、M
B6	低於 1875 元	高於 12 年	X	B、C、D、E、F、 H、I、J、M
A6	低於 1875 元	低於 12 年	X	E、I、M
B6	低於 1875 元	低於 12 年	X	不是 E、I、M

資料來源：本研究整理

表 6 分類結果(維修金額< 465 元)

ItemName (維修零件名稱)	SaleTotal (維修金額)	CarYear (車齡)	LastMove (里程數)	CarFactor (廠牌)
A7	高於 325 元	X	X	X
B3	低於 325 元	X	X	G、K、Others
B3	225~325 元	X	X	B、C、E、F、 M
B9	225~325 元	X	X	不是 G、K、 Others
B8	低於 225 元	X	X	不是 B、C、E、 F、M

資料來源：本研究整理

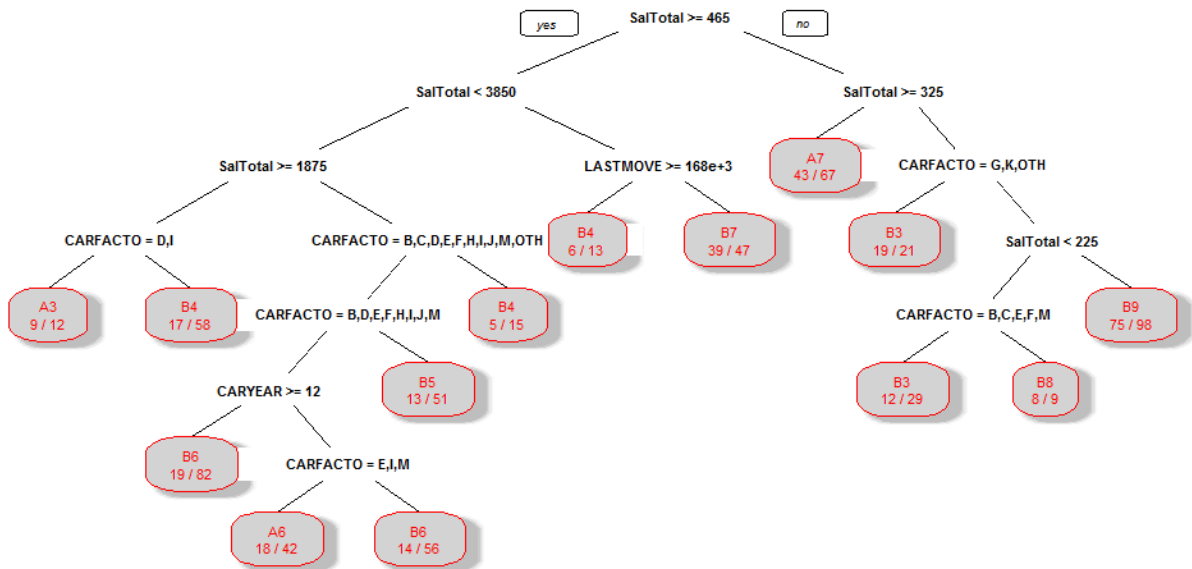


圖 13：R 語言決策樹

資料來源：本研究整理

依照欄位特性與出現與否，討論出分析規則如下：

- ◆ 當維修金額高於 465 元結論
 - 進廠維修汽車里程數高於 16.8 萬時，較常更換 B4 (避震器)，平均維修金額高於 3850 元
 - 進廠做 B7 (汽車保固) 佔多數比例，平均金額高於 3850 元
 - 廠牌 D、I，較常維修 A3 (引擎腳)，平均金額於 1875 ~ 3850 元，與 B4 (避震器) 平均金額高於 3850 元
 - 廠牌 B、C、D、E、F、H、I、J、M、Others，較常維修 B4 (避震器)，平均金額高於 1875 元
 - 廠牌 B、C、D、E、F、H、I、J、M，較常維修 B5 (離合器)，平均金額高於 1875 元；車齡高於 12 年時，較常維修 B6 (煞車來令片) 且平均金額低於 1875 元
 - 當廠牌 E、I、M 車齡低於 12 年時，較常維修 A6 (火星塞)，平均金額低於 1875 元；車齡高於 12 年時，較常維修 B6 (煞車來令片)，平均金額高於 1875 元
- ◆ 當維修金額低於 465 元結論
 - 進廠更換 A7 (外部配件皮帶) 頻繁且

平均維修金額於 325 ~ 465 元

- 廠牌 G、K、Others，較常更換 B3 (機油濾清器)，平均金額低於 325 元，與更換 B9 (空氣濾清器)，平均維修金額於 325 ~ 465 元
- 廠牌 B、C、E、F、M，較常更換 B3 (機油濾清器)，平均金額於 225 ~ 325 元，與更換零件 B8 (油封)，平均維修金額於 225 ~ 465 元

4.4 Weka 迴歸分析

首先使用 Weka 裡的 Preprocess 功能，將樣本資料 test1.csv 做預處理，資料共有 5 個欄位，包括 ItemName (維修零件名稱)、SaleTotal (維修金額)、CarYear (車齡)、LastMove (里程數)、CarFactor (廠牌)。

資料裡的維修金額、車齡與里程數是連續行資料，在此我們將此三個項目運用離散化工具轉換，函數為下列所示：

- filters.unsupervised.attribute.Discretize

之後，選擇 CarYear (車齡) 與 LastMove (里程數) 兩個欄位做分析。

使用 Classify (分類法) 功能裡的 Logistic 做迴歸分析。在此選擇訓練樣本的屬性設定為 CarYear (車齡)，過程採用 Use training set，參數設定如圖 14。

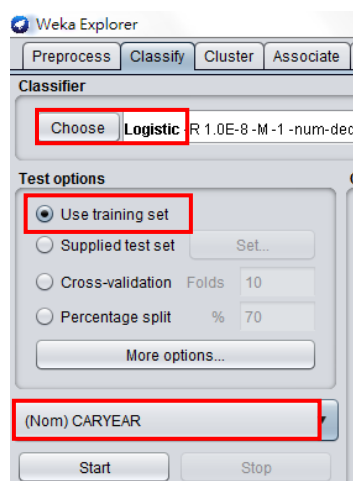


圖 14：Weka 迴歸參數設定

資料來源：本研究整理

之後 Weka 將計算出迴歸分析正確率如圖 15，可看出正確率僅有 42%。

```

=== Summary ===

```

Correctly Classified Instances	251	41.8333 %
Incorrectly Classified Instances	349	58.1667 %
Kappa statistic	0.2131	
Mean absolute error	0.139	
Root mean squared error	0.2636	
Relative absolute error	85.3767 %	
Root relative squared error	92.4795 %	
Total Number of Instances	600	

圖 15：Weka 迴歸分析正確率

資料來源：本研究整理

4.5 Weka 決策樹

接著使用 Classify (分類法) 功能裡的 C4.5-base 演算法 J48 做決策樹分析。

過程是採用訓練及測試方式推導，在此參數設定 70 % 的資料進行訓練，剩餘的 30 % 資料進行測試，在此選擇訓練樣本的屬性設定為 SaleTotal (銷售金額)，參數設定如圖 16。

之後 Weka 將計算出測試資料的正確率如圖 17 所示，可看出分類資料正確率為 74 %。

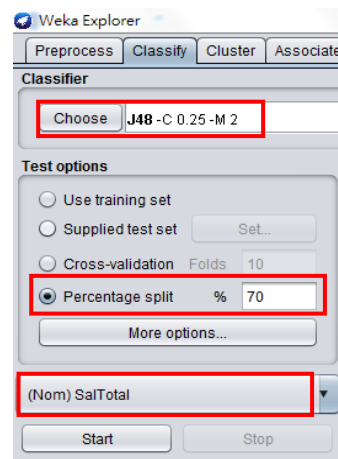


圖 16：Weka 分類參數設定

資料來源：本研究整理

```

=== Summary ===

```

Correctly Classified Instances	133	73.8889 %
Incorrectly Classified Instances	47	26.1111 %
Kappa statistic	0.5078	
Mean absolute error	0.0629	
Root mean squared error	0.1937	
Relative absolute error	53.8457 %	
Root relative squared error	80.0147 %	
Total Number of Instances	180	

圖 17：Weka 分類資料正確率

資料來源：本研究整理

5. 結論

本研究使用 R 語言與 Weka 建立模型並進行交叉比對，結果發現做迴歸分析時使用 Weka 分析的結果正確率僅有 42%，而使用 R 分析的結果迴歸模式達成有效性，車齡與里程數存在著關聯，大部分車齡越高里程數也越高。

做決策樹時使用 Weka 分析的結果發現分類資料正確率為 74 %，而使用 R 分析的結果發現，訓練樣本預測正確率為 81%，相較之下比 Weka 分析的結果高一些。

往後嘗試加入更多資料探勘方法，如類神經網路與關聯規則，期望找出各主要維修項目於何時需做更換，供維修廠做參考依據。

6. 致謝

感謝科技部專題研究經費補助（計畫編號：MOST 105-2410-H-168-002）

參考文獻

- [1] 石油情報出版社 (available online at <http://www.oil.net.tw/>).
- [2] 台北市汽車保養商業同業公會 (available online at <http://www.car-service.org.tw/>).
- [3] 呂文吉，民 104，應用決策樹資料探勘模式於醫療院所資訊設備故障排除之研究—以北部某區域醫院為例，元培醫事科技大學資訊管理系數位創新管理碩士班
- [4] 柯沛丞，民 101，服務品質與顧客信任關係之研究-以台灣汽車維修保養廠為例，國立彰化師範大學工業教育與技術學系
- [5] 范睿昀，民 103，應用資料探勘技術於資源配置預測之研究-以某電腦代工支援單位為例，國立中央大學資訊管理學系在職專班
- [6] 高詮惟，民 104，資料探勘應用於捷運房地產分群與預測，國立政治大學資訊管理系
- [7] 資訊管理 e 化企業的核心競爭能力
- [8] 資料採礦，顧客關係管理與電子行銷之應用，彭文正翻譯
- [9] 資料採掘與 OLAP 理論與實務，林傑斌、劉明德、陳湘編著
- [10] 魏崢，民 99，應用資料探勘技術於汽車維修業之研究，國立虎尾科技大學工業工程與管理研究所
- [11] Berry, M.J.and Linoff, G (1997) . Data Mining Techniques for Marketing , Sale , and Customer Support , Jonh Wiley & Sons , Inc,New York , USA
- [12] Berson,A.,Thearling,K.&Smith (1999) .Building Data Mining Application for CRM, New York:McGraw-Hill.
- [13] M. S. Chen, J. Han, and P. S. Yu, “Data Mining: An Overview from a Database Perspective,” IEEE Transactions on Knowledge and Data Engineering, 1996, pp.866-883.
- [14] R. Agrawal and R. Srikant, “Fast Algorithms for Mining Association Rules,” Proceedings of the 20th International Conference on Very Large Databases, Santiago. Chile, September 1994, pp.487-499.