國立臺東專科學校
National Taitung Junior College

物流用車隊油耗資料分析

指導教授:廖強棋
學生:廖侃紘、邱士桀、李俊齊

摘要

我們所使用的技術是一種專門進行資料探勘的軟體，名叫Weka，在Weka中有各種的功能協助於資料探勘，在這專題中我們採用其中一些功能，所採用的功能為五種，關聯規則、決策樹、分群、貝氏分類及類神經網路，透過這些功能將我們所得到的物流用車隊油耗資料進行分析。

壹、研究動機

由於近年來物流量變得越加龐大使得物流用車的油耗大幅增加，油耗越多則排放的廢氣越多，其廢氣不只對空氣造成嚴重的污染，含有的成分對人體也是一種一級致癌物。所以我們收集了資料使用Weka來作分析，希望能善用這些資料來控制汙染排放。藉由得到的這些物流用車油耗資料，先是透過統計資料後，分別使用關聯規則、分群、決策樹以及類神經網路分析那些平均燃油效率(油耗)、平均總行駛里程和平均總用油量的物流用車在污染上的影響，幫助物流公司理解各類物流用車油耗資料以更好應對不同狀況，有效控制汙染排放。

貳、分析運用

分群分析及關連規則可以對此物流用車各群的數值及關聯後的資料進行比較及猜測，再用總油耗及排氣量作主屬性進行分析比較。決策樹枝幹會顯示分枝節點與節點間的關係，所以決策樹分析可得知這些物流用車在總用油量及排氣量的關聯性。類神經網路是一種模仿生物神經網路的結構和功能的數學模型或計算模型，此資料的類神經網路達到100%。

參、結果分析

一、分群

(圖一)以第零群開始，8月才開始有數據，總用油量平均4442.45公升但排氣量平均有4848 c.c.，再來看第一群，第一群總用油量平均是除了第四群以外數值最大的，而排除第四群的原因是因為在前面也提到關於第四群的數值差距以及對該群的猜測，然而第一群總油耗平均是最大數值，但排氣量卻只有3937 c.c.，是所有群裡排氣量最小，接著是第二群以及第三群，第三群的總油耗平均比第二群的高但排氣量則是第二群較高。而為什麼數值落差會如此多，或許跟分群的計算方式有關，分群的數值計算都是以平均值為主，也就是說該群所擁有的資料筆數以及該群的數值會影響平均值，所以這也是為什麼數值差會這麼多。

Attribute	Full Data (111.0)	0 (20.0)	1 (38.0)	2 (10.0)	3 (29.0)	4 (14.0)
總重(噸)	6.0294	5.8915	4.4397	4.992	5.5628	12.2486
運商類型	市區運輸-多溫	市區運輸-單溫	市區運輸-多溫	市區運輸-單溫	市區運輸-常溫	市區運輸-多溫
排氣量 (c.c.)	4589.4775	4848.45	3937.9211	4487.8	4328.6207	6601
污染管制期別	5.1712	5.95	5	5	5	5
(111年1月)行駛里程 (km)	4674.4505	0	5107.8947	4117	3812.4483	12359.5
(111年1月)用油量 (L)	1050.6599	0	1033.1316	1008.601	823.88	3098.98
(111年1月)燃油效率 (km/L)	3.8319	0	4.8492	3.849	5.0138	4.0843
(111年2月)行駛里程 (km)	4067.6577	0	4566.1316	3144.5	3176.5172	11030.9286
(111年2月)用油量 (L)	887.991	0	942.7105	721.6	652.4138	2614.8571
(111年2月)燃油效率 (km/L)	3.7516	0	4.8645	3.123	4.8552	4.2536
(111年3月)行駛里程 (km)	5338.4324	0	5583.2895	8651.8	4026.3103	12651.4286
(111年3月)用油量 (L)	1015.4054	0	1064.0789	677.6	799.5862	3022.2143
(111年3月)燃油效率 (km/L)	4.2925	0	4.9597	8.005	5.1131	4.2621
(111年4月)行駛里程 (km)	4588.6757	0	5291.2895	2357.8	3924.8966	12205.2857
(111年4月)用油量 (L)	987.0691	0	1015.2895	652.956	787.9686	2971.6443
(111年4月)燃油效率 (km/L)	3.7779	0	4.8371	2.635	5.1793	4.2136
(111年5月)行駛里程 (km)	4748.973	0	5445.7632	2175.4	4224.2759	12567.0714
(111年5月)用油量 (L)	1020.4595	0	1070.3684	561.6	827.9655	3069.2857
(111年5月)燃油效率 (km/L)	3.7887	0	4.8876	2.985	5.0445	4.1914
(111年6月)行駛里程 (km)	4888.2703	0	6070.7368	2459.3	4253.8966	11711
(111年6月)用油量 (L)	1003.3243	0	1174.2368	644.3	734.3793	2786.2857
(111年6月)燃油效率 (km/L)	4.1809	0	5.1629	3.095	6.0031	4.4893
(111年7月)行駛里程 (km)	4937.6306	0	6326.8421	2411.7	4179.069	11596.2143
(111年7月)用油量 (L)	1103.5911	0	1281.2368	664.591	849.2514	3038.3864
(111年7月)燃油效率 (km/L)	3.7377	0	4.85	2.516	5.1417	4.0221
(111年8月)行駛里程 (km)	4804.7928	1274.1	5669.6053	1401.5	4092.5862	11407.5
(111年8月)用油量 (L)	998.955	272.65	1104.9474	450.4	730.9655	2695.7857
(111年8月)燃油效率 (km/L)	4.3789	1.727	5.0311	2.254	5.9534	4.6536
(111年9月)行駛里程 (km)	4712.955	3060.3	4921.6579	583.7	4032.1379	10867.1429
(111年9月)用油量 (L)	1041.2613	661.6	985.1053	217.7	815.2759	2792.4286
(111年9月)燃油效率 (km/L)	4.055	3.2605	4.6895	1.189	4.76	4.0543
(111年10月)行駛里程 (km)	5104.8018	4126.4	5135.4211	739	4222.6897	11365.0714
(111年10月)用油量 (L)	1117.4469	915.574	1009.7105	243.021	840.6266	2896.2679
(111年10月)燃油效率 (km/L)	4.1942	3.8145	4.8203	1.525	4.5931	4.1179
(111年11月)行駛里程 (km)	4229.0721	5182.85	2803.1053	877.2	3143.1034	11380.7143
(111年11月)用油量 (L)	967.2342	1198	562.2895	239.9	654.0345	2905
(111年11月)燃油效率 (km/L)	3.6514	4.006	3.6611	1.541	3.9079	4.0943
(111年12月)行駛里程 (km)	3987.9369	5979.45	2379.1053	722.2	2877.2069	10143.2143
(111年12月)用油量 (L)	908.6289	1394.778	459.3629	194.139	605.4869	2571.8536
(111年12月)燃油效率 (km/L)	3.3705	4.3565	3.3189	1.106	3.1793	4.115
總行駛里程 (公里)	56083.6486	19623.1	59300.8421	29641.1	45965.1379	139285.0714
總用油量 (公升)	12102.1081	4442.45	11702.6053	6276.5	9122.0345	34463
平均燃油效率 (km/L)	4.8507	4.4195	5.0084	4.846	5.2503	4.2143

圖一、分群結果

國立臺東專科學校
National Taitung Junior College

物流用車隊油耗資料分析

指導教授:廖強棋
學生:廖倀絟、邱士桀、李俊齊

參、結果分析(續)

二、關聯規則

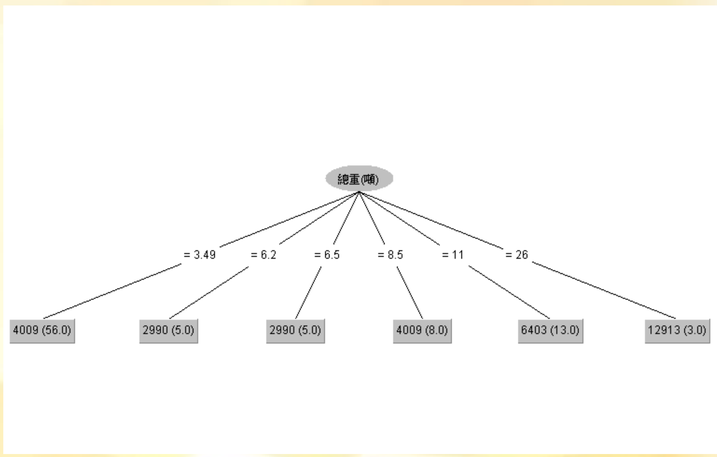
1.以下的關聯規則(圖二)都是總重和行駛里程去關連到排氣量，主要我們發現重量坐落在3.49噸的物流用車其排氣量均為4009c.c.，且在3~7月該類型的物流用車的行駛里程都坐落在3000多~4000多附近。在最後一項關聯規則我們能夠發現他的運務類型也跟總重和排氣量有關，而市區運輸-常溫的物流用車總重多數坐落在3.49噸，排氣量坐落在4009c.c.。

2.而以下的關聯規則(圖三)我們發現前兩項跟也前面的總重對應到排氣量一樣，大多總重為3.49的物流用車排氣輛為4099c.c.，而11噸的物流用車排氣量為6403c.c.，接著是依照行駛里程的關聯，我們的行駛里程(km)/1000在3~2的區間汙染管制期別都在5，而汙染管制期別=5表示該車是在101年1月1日 至108年8月31日出場或進口之年月。

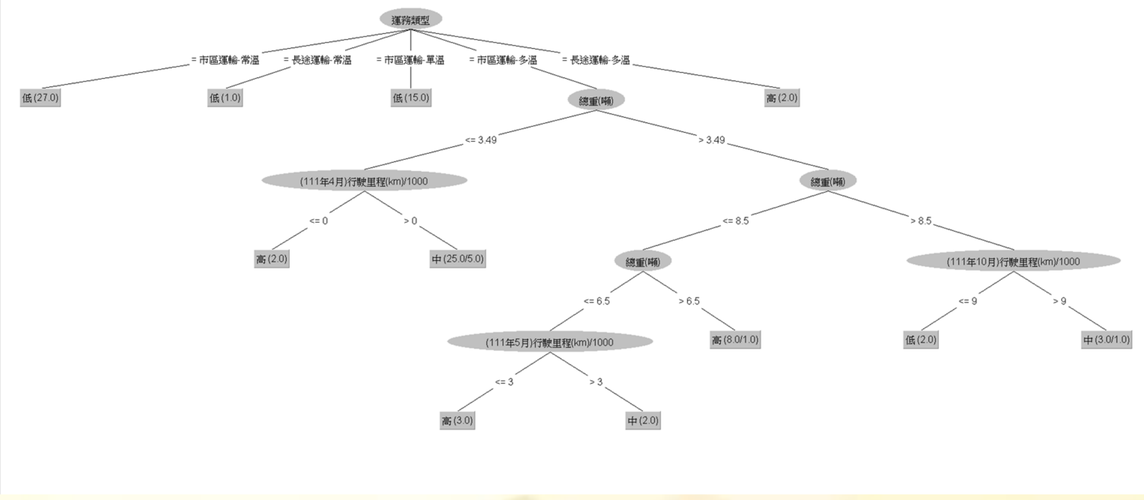
三、決策樹

1.此決策樹(圖四)使用了J48演算法，先以[排氣量]作為屬性來搭建決策樹，[總重]=3.49為[排氣量]4009有56筆、[總重]=6.2為[排氣量]2990有5筆、[總重]=6.5為[排氣量]2990有5筆、[總重]=8.5為[排氣量]4009有8筆、[總重]=11為[排氣量]6403有13筆、[總重]=26為[排氣量]12913有3筆。

2.此決策樹(圖五)以[總用油量]作為屬性，首先運務類型分出了[市區運輸-常溫]為低有27筆、[長途運輸]為低有1筆、[市區運輸-單溫]為低有15筆、[長途運輸-多溫]為高有2筆，其中[市區運輸-多溫]能以[總重(噸)]往下細分。[總重(噸)]小於等於3.49噸為[(111年4月)行駛里程(km)/1000]，但往下細分後出現了0(資料空缺所以用0來代替)，所以這部份的分支不怎麼重要。而大於3.49噸又能以同樣的屬性[總重(噸)]進行細分，大於8.5噸能分到[(111年10月)行駛里程(km)/1000]，小於等於9km/1000為低有2筆、大於9km/1000為高有3筆。小於等於8.5噸還能再度細分，大於6.5噸為高有8筆，則小於等於6.5噸可分到屬性[(111年5月)行駛里程(km)/1000]，大於3km/1000為中有2筆、小於等於3km/1000為高有3筆。



圖四、決策樹結果(一)



圖五、決策樹結果(二)

```
1. 總重(噸)=3.49 (111年3月)行駛里程(km)/1000=3 9 ==> 排氣量(c.c.)=4009 9   conf:(1) < lift:(1.41)> lev:(0.03) [2] conv:(2.6)

2. 總重(噸)=3.49 (111年5月)行駛里程(km)/1000=3 9 ==> 排氣量(c.c.)=4009 9   conf:(1) < lift:(1.41)> lev:(0.03) [2] conv:(2.6)

3. 總重(噸)=3.49 (111年5月)行駛里程(km)/1000=4 9 ==> 排氣量(c.c.)=4009 9   conf:(1) < lift:(1.41)> lev:(0.03) [2] conv:(2.6)

4. 總重(噸)=3.49 (111年6月)行駛里程(km)/1000=2 9 ==> 排氣量(c.c.)=4009 9   conf:(1) < lift:(1.41)> lev:(0.03) [2] conv:(2.6)

5. 總重(噸)=3.49 (111年7月)行駛里程(km)/1000=4 9 ==> 排氣量(c.c.)=4009 9   conf:(1) < lift:(1.41)> lev:(0.03) [2] conv:(2.6)

6. 運務類型=市區運輸-常溫 27 ==> 總重(噸)=3.49 排氣量(c.c.)=4009 14   conf:(0.52) < lift:(0.83)> lev:(-0.03) [-2] conv:(0.73)
```

圖二、關聯規則結果(一)

```
1.總重 (噸)=3.49  74 ==> 排氣量 (c.c.)=4009  74  conf:(1) < lift:(1.35)> lev:(0.17) [19] conv:(19.33)

2. 總重(噸)=11 13 ==> 排氣量 (c.c.)=6403 13   conf:(1) < lift:(6.92)> lev:(0.12) [11] conv:(11.12)

3. (111年1月) 行駛里程 (km) / 1000=2  16 ==>  汙染管制期別=5  16 <conf:(1) > lift:(1)  lev:(0) [0] conv:(0)

4.(111年2月) 行駛里程 (km) / 1000=3  16 ==>  汙染管制期別=5  16 <conf:(1) > lift:(1)  lev:(0) [0] conv:(0)

5.(111年3月) 行駛里程 (km) / 1000=3  16 ==>  汙染管制期別=5  16 <conf:(1) > lift:(1)  lev:(0) [0] conv:(0)

6.運務類型=市區運輸-常溫  30 ==> 汙染管制期別=5  28  <conf:(0.93) > lift:(1.13)  lev:(0.03) [3] conv:(1.71)
```

圖三、關聯規則結果(二)

四、貝氏分類

$$p(H|E) = \frac{p(E|H) \times p(H)}{p(E|H) \times p(H) + p(E|\neg H) \times p(\neg H)}$$

$$p(H_i|E_1E_2 \dots E_n) = \frac{p(E_1E_2 \dots E_n|H_i) \times p(H_i)}{\sum_{k=1}^{m} p(E_1E_2 \dots E_n|H_k) \times p(H_k)}$$

$$p(H_i|E_1E_2 \dots E_n) = \frac{p(E_1|H_i) \times p(E_2|H_i) \times \dots \times p(E_n|H_i) \times p(H_i)}{\sum_{k=1}^{m} p(E_1|H_k) \times p(E_2|H_k) \times \dots \times p(E_n|H_k) \times p(H_k)}$$

$$p(H_i|E) = \frac{p(E|H_i) \times p(H_i)}{\sum_{k=1}^{m} p(E|H_k) \times p(H_k)}$$

圖六、貝氏分類之假設及證據計算機率之公式

```
=== Summary ===

Correctly Classified Instances          83               92.2222 %
Incorrectly Classified Instances         7                7.7778 %
Kappa statistic                          0.7719
Mean absolute error                      0.0775
Root mean squared error                  0.278
Relative absolute error                 24.8754 %
Root relative squared error             70.8684 %
Total Number of Instances               90

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                 0.918     0.059     0.985     0.918     0.950      0.782     0.986     0.997     低
                 0.941     0.082     0.727     0.941     0.821      0.782     0.979     0.863     高
Weighted Avg.    0.922     0.063     0.937     0.922     0.926      0.782     0.985     0.972

=== Confusion Matrix ===

  a  b   <-- classified as
 67  6 |  a = 低
  1 16 |  b = 高
```

圖七、貝式分類結果

TP=16

FP=6

TN=67

FN=1

True Positive Rate（真正率，TPR）或靈敏度（sensitivity）

TPR = TP /（TP + FN）=16/(16+1)=0.9411=94.11%

正樣本預測結果數 / 正樣本實際數

True Negative Rate（真負率，TNR）或特指度（specificity）

TNR = TN /（TN + FP）=67/(67+6)=0.9178=91.78%

負樣本預測結果數 / 負樣本實際數

False Positive Rate（假正率, FPR）

FPR = FP /（FP + TN）=6/(6+67)=0.0821=8.21%

被預測為正的負樣本結果數 /負樣本實際數

False Negative Rate（假負率，FNR）

FNR = FN /（TP + FN）=1/(16+1)=0.0588=5.88%

被預測為負的正樣本結果數 / 正樣本實際數

準確率（Accuracy）= (tp+tn)/(tp+fp+fn+tn)=(16+67)/(16+6+1+67)=0.9222=92.22%

精確率（Precision）= tp/(tp+fp)=16/(16+6)=0.7272=72.72%

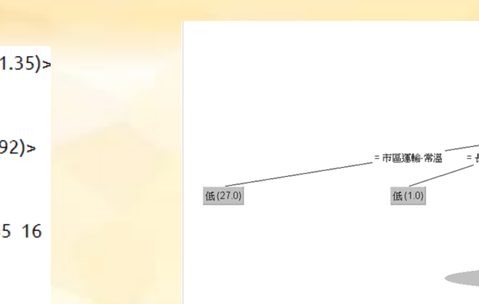
召回率（Recall）= tp/(tp+fn)=16/(16+1)=0.9411=94.11%

F1-score = 2 / ((1/ Precision) + (1/ Recall))= 2 * Precision * Recall / (Precision + Recall)=2* (16/(16+6)) * (16/(16+1)) / ((16/(16+6)) + (16/(16+1)))=0.8205=82.05%

圖八、Confusion Matrix、TP、TN、FP、FN、Precision、Recall及F1-score之計算

五、類神經網路

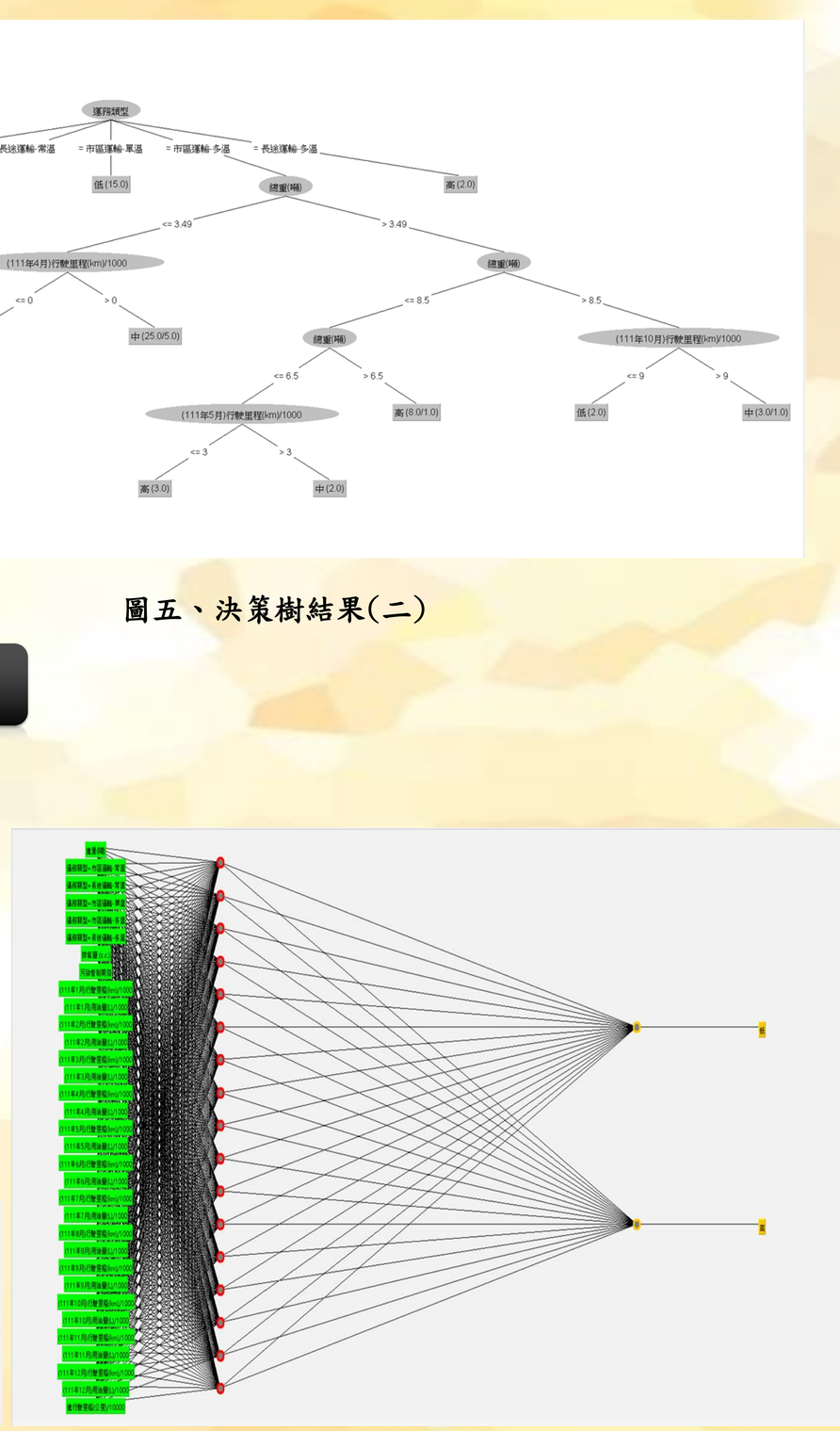
此圖取自於《人工智慧：智慧型系統導論》類神經網路篇中。



圖九、簡單的類神經網路圖

```
=== Summary ===

Correctly Classified Instances          18              100      %
Incorrectly Classified Instances         0                0      %
Kappa statistic                          1
Mean absolute error                      0.0009
Root mean squared error                  0.0022
Relative absolute error                  0.2834 %
Root relative squared error              0.5356 %
Total Number of Instances               18

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000     0.000     1.000     1.000     1.000      1.000     1.000     1.000     低
                 1.000     0.000     1.000     1.000     1.000      1.000     1.000     1.000     高
Weighted Avg.    1.000     0.000     1.000     1.000     1.000      1.000     1.000     1.000

=== Confusion Matrix ===

  a  b   <-- classified as
 14  0 |  a = 低
  0  4 |  b = 高
```

圖十、類神經網路結果



圖十一、類神經網路圖

致謝

感謝財團法人車輛研究測試中心(ARTC)提供實際資料供我們研究分析，
財團法人車輛研究測試中心(ARTC)官網:https://www.artc.org.tw
與QR-code(如右圖)。