

## 支撐向量機分類器於商業型用戶違章用電之應用

# The Application of Support Vector Machine Classifier to Commercial Customer Electricity Theft

<sup>1</sup> 卓明遠

Ming-Yuan Cho

<sup>1</sup> 鄭淵澤

Yuen-Tse Cheng

<sup>2</sup> 王念中

Nien-Chung Wang

<sup>2</sup> 黃佳文

Chia-Wen Huang

<sup>1</sup> 國立高雄應用科技大學

台灣 高雄市

Department of Electrical Engineering

National Kaohsiung University of Applied Sciences

Kaohsiung, TAIWAN

mycho@ma

il.ee.kuas.edu.tw

Lucky731214@

hotmail.com

<sup>2</sup> 台灣電力綜合研究所

台灣 台北市

Power Research Institute

Taiwan Power Company

Taipei, TAIWAN

u629802@taipower.com.tw

### 摘要

本論文提出一種支撐向量機 (Support Vector Machine, SVM) 網路之模型辨識技術, 藉由用戶合理用電模型之建立與比對, 進行商業型用戶違章用電之分類。首先, 本文收集台灣代表性之商業型用戶用電歷史資料, 推導其夏月與非夏月之合理用電模型, 再者應用支撐向量機網路模型, 針對商業型用戶違章用電行為與特性, 建立商業型用戶之違章用電分類器, 針對各類型用戶進行資料分類與辨識, 推求用戶違章用電之模式, 並找出有違章用電可能之用戶及其違章用電之度數。

關鍵字: 商業型用戶違章用電模式、模式辨識技術、支撐向量機

### Abstract

This thesis proposes support vector machine based pattern recognition technique to classify commercial customer electricity theft by establishing and comparing with the various rational load patterns. First, in this thesis, Taiwan commercial customers' historical electricity data is collected to derive the summer and non-summer reasonable power consumption model. Moreover, the SVM network model is employed to train the selected commercial customer data set to establish the commercial customer electricity theft classifier and then the electricity theft electricity KWH can be derived by analyzing and recognizing historical data in database.

Key words: Commercial customer electricity theft model, Pattern recognition technique, Support vector machine

### I. 緒論

用戶違章用電 (竊電) 多年來一直存在於配電系統, 台電各區營業處亦針對各種可能之竊電方式分析, 並試圖開發零序電流檢知自動發報器、安裝 KV 表等方式解決問題, 但因現場易受破壞及無實際之長期違章用電資訊佐證, 外加現場執行任務之困難度與危險性, 致使用戶違章用電問題迄今仍無法獲得有效解決。

本論文的研究目的在於應用台電最適契約容量資料庫及商業用戶週期每月用電資料模型辨識技巧, 建立以 SVM 為基礎之違章用電分析系統, 藉由台電最適契約容量資料庫裡的每月用電資料為基礎建立並統計於合理用電模型曲線來進行比較, 應用模型辨識技巧找出有可疑性之用電用戶。而台電最適契約容量資料庫備有從 2007 年至今的歷史每月用電資料, 資料充足以作為往後執行取締或求償之依據, 並有效嚇阻違章用電之行為。

### II. 用戶用電分析

用戶分類及統計

本論文之研究用戶篩選是選擇用電量甚大之服務業部門, 依經濟部能源委員會 2009 年能源統計[1], 台灣地區總電力消費量 229067.6 百萬度電, 尖峰負載 33858000kW, 而服務業部門電力消費量 45057.6 百萬度, 約佔總電力消費量之 19.67%, 由此可得知其耗電量之大。

一般俗稱的零售通路, 大概區分為: 便利商店、超級市場及量販店。然而根據全台灣各區處綜合商品零售業的不同而分別取得一些合理用電模型, 從最適契約

容量資料庫得知綜合商品零售業之連鎖便利商店、量販店及超級市場在全台灣各區處裡佔用之商業用電的百分比都相當高，由此可知這些業別如進行竊電行為，其用電度數是相當可觀的。

### 1. 台灣地區量販店店家數分布：

由圖 1 顯示在量販業由法商家樂福與統一集團所成立之家樂福量販店為排名為第一，佔全部量販業者百分比為 42%，其總店家數為 63 家，其次為潤泰集團所成立之大潤發量販，總店家數以 24 家排名為第二，特力集團及翠豐集團所成立 B & Q 特力屋其店家數為 22 家排名第三，由遠東集團投資經營之愛買吉安量販店其店家數為 17 家排名為第四，其他則是少量數販業全部加總大約為 14 家。

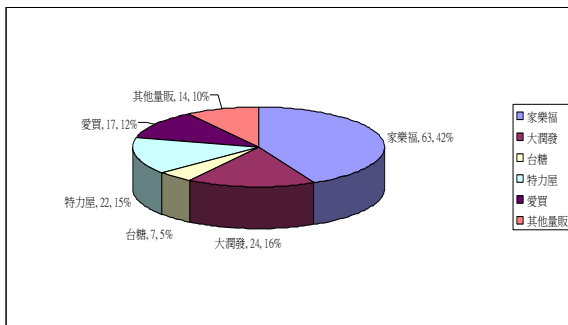


圖 1 台灣國內連鎖量販店數之分佈圖

### 用戶行業類別及負載特性分析

一般各量販店規模頗大，平均賣場面積 15,462m<sup>2</sup>、約 5,000 坪，若加上停車位則為 25,875m<sup>2</sup>、約 8,000 坪，空調、照明用電設備幾乎每天運轉約 13 小時，其中冷凍冷藏用電設備幾乎全天運轉，由於空間廣大須以較長的風管送風到賣場之各部位，其耗能也是相當可觀的，而大多數量販店用電需量為超過 1,000 kW 之能源大用戶。 [2]

### 用戶負載組成與負載用電分析

本論文針對全台灣各區處之用電密度最高的零售業來做分析，所以選擇了連鎖便利商店、量販店、超級市場三種零售通路來做為分析，而一般連鎖便利商店跟台電訂定契約容量通常大約是 30kW，而由於量販店大小不同所以其負載量也相差甚大，根據查證其跟台電簽約之契約容量約為 800kW~2200kW，範圍差距大，所以本論文取以知經濟部統計之 6 家之平均契約為 1,221kW 當為參考負載組成模型在與測得 10 家量販店

用電度數進行比對與估算，來做為本論文之合理用電曲線，至於超級市場，以探討排名之龍頭超級市場其該市場跟台電訂定契約容量通常大約為 90kW 並以實地訪查略估計算其合理負載之模型來取樣比對。表 1 說明了量販店負載組成及各個負載啟閉用電時間。

表 1 量販店負載組成及各個負載啟閉用電時間

量販店用電		
坪數	7156 坪	耗能
種類	量販店	
負載組成	空調設備	550 kW
	冷凍、冷藏設備	208 kW
	照明設備	354 kW
	其他設備	147 kW
用電時間	空調設備 9:00~22:00	/
	照明設備 9:00~22:00	
	其他設備 9:00~22:00	
	冷凍、冷藏設備為 24 小時用電	

### 用戶違章用電類型之分析

所謂違章用電是指改裝、破壞電表和使電表計度變慢，以降低電費[3,4,5]。為了防止此類事件發生，除了透過民眾檢舉或電力公司稽查人員巡視查獲外，也可以透過觀察用戶之用電量是否出現異常變化的可疑用電資料中去深入追查。國內違章用電類型大致上可以分為以下幾種方法：

方法一：加裝迴路。經過用戶改造電表將接線繞表外以後，電流就會繞越電表，直接進去屋內，不必經過電表加以計算度數。

方法二：回撥度數。有些用戶會把電表上的鉛封破壞掉，然後將電表打開，電表上面計算度數的數字就可以隨用戶撥動。

方法三：遙控啟動。有些用戶會在電表內加裝遙控裝置，控制電表運轉，只要遙控一啟動，電表就停止不動。

方法四：更換內部齒輪(以機械式電表為主)。有些用戶會打開電表，直接更換電表內部零件，改為裝設偽造品，如劣質齒輪。

方法五：使用強力磁鐵。還有些用戶會利用強力磁鐵的吸力，將電表的轉盤吸住或是減慢其速度，影響電表計量之速度。

方法六：使用並聯電容器。適當的並聯電容器能減低線路阻抗並能改善功率因數，但是當電容器過度補償，反而會造成更差的功率因數。

### III. 支撐向量機理論

SVM是由Vapnik等人根據統計學理論中的結構風險最小化(structural risk minimization, SRM)原理所發展出來的理論[6]。SVM的基礎主要是在討論二元分類問題，在高維度特徵空間中尋找一個超平面作為二類的分割(如圖2所示)，以保證最小的分類錯誤率[7]。從幾何上說，支撐向量機就是要在  $n$  維空間中尋找最佳決策面，該決策面能將正例和反例作最好的區分，並使正例和反例之間的分類間隔為最大。

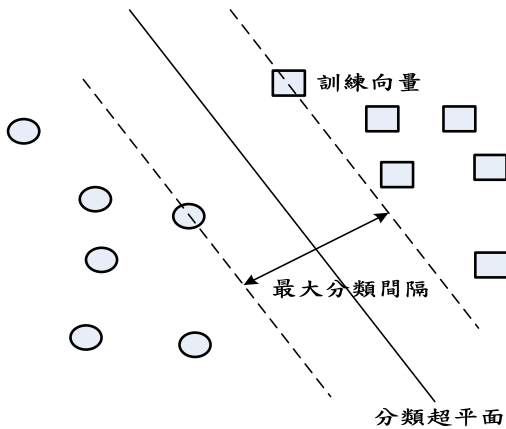


圖2 線性可分之分類超平面

SVM 針對一般在低維度空間線性可分類之問題，皆能將不同類型之樣本分離，若是遇上非線性不可分類之問題時，SVM 則可以經由非線性映射轉化(Feature Mapping:  $\Phi: x \rightarrow \varphi(x)$ )為某個高維度特徵空間的線性問題，而在高維度特徵空間中會對應到低維度輸入空間的一個曲面上，再利用核函數(Kernel Function)將不可分類的問題轉變為線性可分類的問題，並在變換空間中尋找最佳的分類面。

從圖 3 可簡述線性分類的概念，當有一群的資料可以利用直線將資料區分成兩類，此直線的方程式為  $(w \cdot x) + b = 0$ ，而 Support Vector 在直線的左邊為一群，在右邊的為一群，依分類公式決定。可將上述不等式的規範形式合併為如下緊湊型式

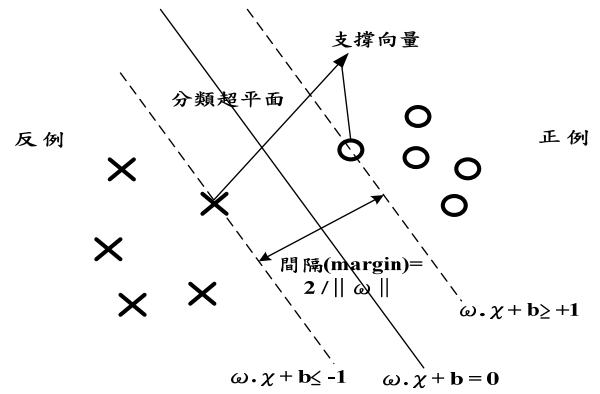


圖 3 支撐向量機分類器架構

$$y_i(w \cdot x_i + b) \geq 1 \quad i=1,2,\dots,l \quad (1)$$

要使分類間隔最大化，就是使  $2/\|w\|$  最大。因此建構最佳分類超平面的問題可轉化為在滿足上式條件下最小化的問題。

$$\Phi(w, b) = \frac{1}{2} w \cdot w$$

$$\xi \geq 0, i=1,2,\dots,l \quad (2)$$

另外，考慮到可能存在一些樣本不能被超平面正確分類，因此引入正值的鬆弛變量 (Slack variable)

$$(3)$$

因此建構廣義最佳分類超平面問題轉化為在

$$y_i(w \cdot x_i + b) \geq 1 - \xi_i \quad i=1,2,\dots,l,\dots \quad (4)$$

得到最小化函數

$$\Phi(w, b) = \frac{1}{2} w \cdot w + C \sum_{i=1}^l \xi_i \quad (5)$$

上式中第二項為經驗(風險)誤差， $C$  為正值懲罰常數 (Regularized constant)， $C$  越大，對錯誤的懲罰越重。其中第 1 項是樣本到超平面的距離要盡量大，從而提高泛化能力；第 2 項要使誤差盡量小。

### IV. 建構用戶違章用電分類系統

理論上用戶的合理用電曲線需要根據相同用戶的用電曲線之歷史資料，經由嚴謹的統計與推估技術堆導出來。本論文之資料庫備有完整的三年的月用電資料，故歷史資料完善，然後再透過訪察及實地探勘，將對不同種類商家用戶的各項用電設備與用電時程列出，然後合成出各種商業之合理負載曲線，其合理負載曲線分析

流程圖如圖 4 所示。

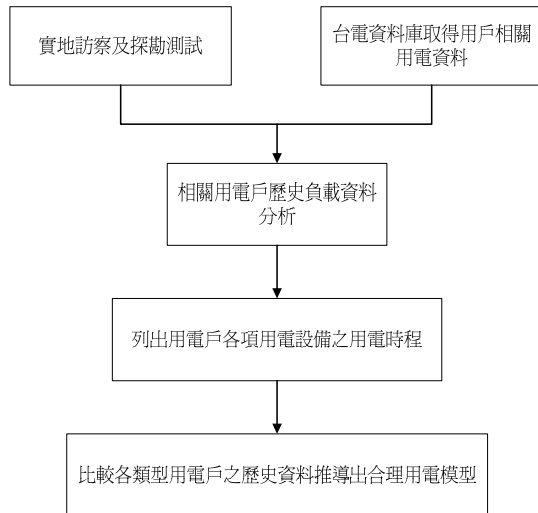


圖 4 合理負載模型建立流程

根據圖 4 推估便利商店、量販店、超級市場日負載合理模型，並同時求解該用戶於某日之總用電量，其數學模式為

$$R(t) = \left( \sum_{n=0}^{c_1} \sum_{m=0}^{c_2} P_n \cdot T \cdot m \right) \quad (6)$$

$$K = \sum_{t=1}^{24} R(t) \quad (7)$$

其中  $R(t)$ ：單日區間小時合理用電度數(kWH)

$K$ ：單日合理用電度數(kWH)

$P$ ：單一負載消耗之實功率(kW)

$n$ ：用戶之負載種類

$m$ ：用戶單一負載之數量

$T$ ：用電時間( $T = 1, 2, 3, \dots, 24$ )

$t$ ：時間常數

$c_1$ 、 $c_2$ ：常數值

### 1. 量販店合理日負載曲線

量販店目前主要營業項目為各類家電、家庭民生產品、各類生鮮食品、服飾、文具等各項產品。由於量販店營業面積大，所以對於空調及照明設備耗電量遠大於其他設備，整年而在耗電尖峰又可分為兩種情形，一為夏季空調使用造成夏月耗電量成長，二為量販店人潮所造成之耗電，例如週年慶與各式活動人潮眾多，造成空調使用與冷凍冷藏的負荷增大，如舉例 8 家量販店一整年用電之情形，其中在夏季用電量與非夏季用電量之差距平均約為在 8 萬度~10 萬度。

而量販店除了冷凍冷藏 24 小時用電外，其餘幾乎都在營業時間才有運轉而營運時間為 9：00~22：00 以表示其負載取其平均，如圖 5 所示。表 2 說明推估合

理量販店在夏月與非夏月月用電量及量販店一年的總用。

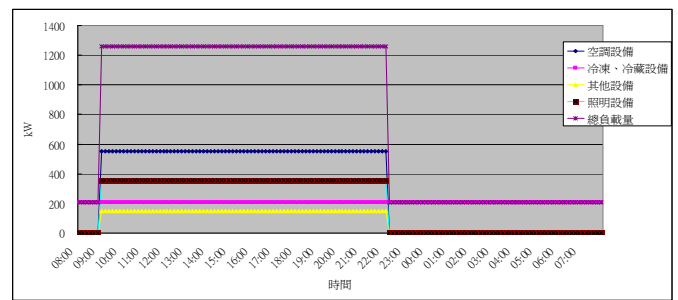


圖 5 量販店合理之負載曲線比較圖

表 2 量販店夏月與非夏月月用電及一整年之用電量

量販店	夏月月用	非夏月月	整年總用
	電度數	用電度數	
	559,650	469,650	5,995,800

一般竊電商店在竊電之習性不會完全竊電，因為避免台電稽查人員發現當月與歷史月用電量差太多，有些商家可能甚至已經沒有經營，或者進行整修。所以本論文之竊電用戶篩選必須經過兩階段的 SVM 分類器作為分類，第一階段先把上述之商店篩選剔除，然後再將第一階段已篩選過後，留下之用戶在進行第二階段篩選出合理用電與不合理用電，預先定義出+1 值為合理用電曲線，-1 值為不正常使用曲線。圖 6 說明瞭整個 SVM 兩階段篩選過程。第一階段篩選是將商店用電週期合理用電度數的 20% 以下篩選過濾出來，因這些或許已經沒有營業或整修中，只剩一些少許設備用電，此列為第一階段合理戶數，接下來第二階段，將第一階段所篩選出來的不合理資料再進行一次的資料辨識，將用戶用電週期合理用電度數的 80% 以上篩選出來，此為合理用戶數，剩下來的戶數具有不合理用電度數之戶數。

本論文輸入數據之整理，為擷取資料庫用戶實際月用電資料，每一個月為一個用電數據，由於商家用電都為一整年用電，所以分別週期都以月為單位，分成 12 個月，分別以 X1~X12 代替 12 個數據，為一筆資料。經由篩選過後，預先定義出+1 值為合理用電曲線，-1 值為不正常使用曲線。利用 NeuroSolutions 中的 SVM 模型建構分類系統之模型架構。首先將需要應用到之文字檔建立完成

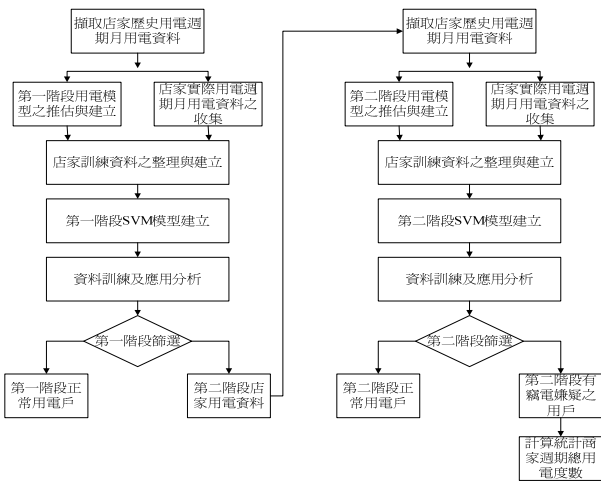


圖 6 SVM 兩階段篩選過程流程圖

此操作介面之主要功能，將查詢用戶條件一一按照步驟選取，到了步驟五選擇不合理戶數統計階段，將已建好 SVM 網路模型連結到資料庫，將資料庫所查詢到之資料丟入 SVM 網路模型做分類，然後將各個查詢到的資料之月用電度數，以用電週期為單位加總起來，並且將總用電度數轉入流動電費，再將戶數加總。圖 7 說明介面操作流程。圖 8 為人機介面顯示圖。

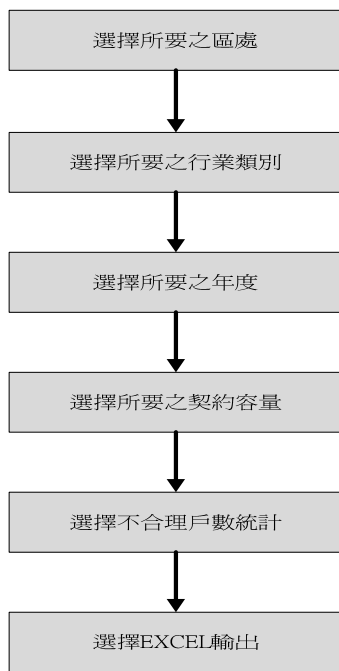


圖 7 介面操作流程



圖 8 人機介面顯示圖

## V. 測試與結果分析

本文利用 ASP.NET 撰寫程式介面將區處資料從資料庫收集整理分析，接著藉由各商業用戶之歷史資料及合理用電模型反覆訓練，找出在低錯誤率及低平均誤差所建立而成之 SVM 分類模型，再將經過兩次的 SVM 篩選之資料輸入支撐向量機模型中，所得到之數據即為我們所要之不合理用電結果。下列依序說明利用兩次 SVM 分類模型篩選，並應用 NeuroSolutions 軟體測試及分類所得之結果。首先選取支撐向量機模型，作為訓練之模型，接著按照第肆章所描述之步驟(包含前置步驟、建立模型、測試及最後之應用)建構以支撐向量機為基礎之用戶曲線分類系統，最後分批將第一次 SVM 分類篩選及第二次 SVM 篩選之月用電資料(西元 2007 至 2009 年)，放入最後之應用測試中，並將測試結果顯示出來。

### 一、量販店 (契約容量:1259kW 用電週期:十二個月)

分別將第一次 SVM 分類模型及第二次 SVM 分類模型之合理負載曲線及不合理負載曲線所建構之用戶曲線分類系統為第一部分，接著將已知目標值之數據輸入進行測試，測試此用戶曲線分類系統之分類是否已合乎需求，並分別顯示測試結果。因為用電週期為十二個月，所以輸入數據從 X1 到 X12 為一筆，其中 X1~X8 為非夏月用電度數而 X9~X12 為夏月用電度數。

第一階段 SVM 分類測試值，分別以 5 筆合理、不合理曲線資料輸入系統中測試，圖 9 說明第一階段 SVM 篩選測試值所示為第一階段測試結果，由圖 9 可以看出，在測試資料中之 desired 值為 5 筆-1(不合理曲線)及 5 筆 1(合理曲線)，而其輸出值可以很明顯的將 desired 值之輸入資料大致分為二類。同樣的步驟和第一階段一模一

樣，圖 10 說明第二階段 SVM 篩選測試值所示為第二階段測試結果，而其輸出值可以很明顯的將 desired 值之輸入資料大致分為二類。很明顯的 desired 值和 out 值很相近，代表著合理用電模型建立成功。

Des y	Out y
1.000000000000	0.827663925944
1.000000000000	0.580069224056
1.000000000000	0.850760665389
1.000000000000	0.823974499179
1.000000000000	0.737973698014
-1.000000000000	-0.813504666634
-1.000000000000	-0.836543354883
-1.000000000000	-0.785401262016
-1.000000000000	-0.660772078071
-1.000000000000	-0.729021862257

圖 9 第一階段 SVM 篩選測試值結果(量販店-1259kW)

Des y	Out y
1.000000000000	0.818952341491
1.000000000000	0.838695877266
1.000000000000	0.698017379171
1.000000000000	0.716316010415
1.000000000000	0.763799868941
-1.000000000000	-0.705860600874
-1.000000000000	-0.773993584459
-1.000000000000	-0.771769080023
-1.000000000000	-0.798699162393
-1.000000000000	-0.778091389474

圖 10 第二階段 SVM 篩選測試值結果(量販店-1259kW)

因為量販店用電週期為十二個月，以我們分別將 2007 年、2008 年和 2009 年之一月至十二月的月用電資料投入合理用電模型中，分析結果如圖 11 說明瞭第一階段與第二階段分析結果之戶數統計

### VI. 結論

本文提出利用支撐向量機所建立之分類器，作為用戶違章用電分類之方法，這個方法在本文中以支撐向量機網路來對用戶用電正常與否做有效的診斷，基於上述之優點，本文在實際運用面，可歸類出此套系統之具體成果，其敘述如下：

1. 本研究已經成功建立三種商業用電戶(便利商店、量販店、超級市場)合理用電模型，辨識效果良好，並藉由人機介面成功辨識出不合理用電戶數及度數。
2. 本研究所建立之系統可支援台灣電力公司作為執行違章用電查緝工作，且所蒐集之用戶月用電資料，有助於提升成案機率。
3. 本系統利用最適契約容量資料庫方式取得用

戶月用電，免除現場資料擷取之困難度，增加資料取得之方便性，以減少和用戶面對面取締之危險性。

最後，本論文以資料庫實際資料(包含 2007 年至 2009 年)進行違章用電分類與辨識分析，由所得之結果可驗證本論文之實用性以及所提 SVM 分類器之效能。

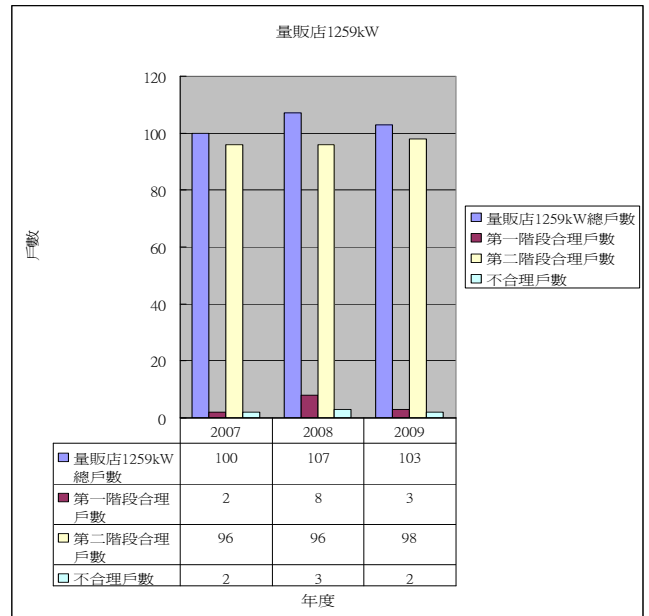


圖 11 第一階段與第二階段分析結果之戶數統計  
(量販店-1259kW)

### 參考文獻

- [1] 2009年台灣能源統計年報：經濟部能源會。
- [2] 量販店訪測服務報告，產業約能源技術服務（經濟部能源研究發展基金九十一年度計畫）。
- [3] C. J. Bandim, J. E. R. Alves Jr, A.V. Pinto Jr, F.C. Souza, M.R.B. Loureiro, C. A. Magalhães and F.Galvez-Durand, “Identification of Energy Theft and Tampered Meters Using a Central Observer Meter: A Mathematical Approach,” *IEEE Transmission and Distribution Conference and Exposition*, Vol.1, pp. 163–168, 2003.
- [4] Saptarshi De, Rahul Anand, A Naveen and Sirat Moinuddin, “e-Metering Solution for checking energy thefts and streamlining revenue collection in INDIA,” *IEEE Transmission and Distribution Conference and Exposition*, Vol.2, pp. 654–658, 2003
- [5] Rahul hand, Saptarshi De, A Naveen, “Design and Development of Vigilant Energy Metering System

(VEMS) and its Applications,” *IEEE Student Conference on Research and Development*, pp. 15–18, 2003.

- [6] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag, New York, 1995
- [7] Kuan-yu Chen, “Application of Support Vector Regression in Forecasting International Tourism Demand,” *Tourism Management Research* Vol. 4 No. 1, pp. 81-97, June, 2004.