

以資料探勘技術分析高速公路服務水準之研究

朱松偉¹ 林佑霖²

摘要

本論文以大量且連續的探針車資料，結合資料探勘之方法與技術，選取國道一號台北圓山交流道至桃園南崁交流道之路段，經過資料篩選與處理流程後，建構出本研究之道路服務水準推估模式，讓本研究可真實地應用於高速公路之管理與控制上。本研究顯示透過長時間探針車資料，不僅可有效區隔研究區內不同時段、不同日期類型的尖峰/離峰時間以及服務水準，計算出各空間單元之服務水準，且速度變化以及尖峰與離峰之趨勢與前人研究成果一致。本研究也成功證明於 VD 缺乏之路段，亦可估算出可靠之平均路段速率，可作為相關單位交通管理與施政之參考。

關鍵詞：探針車、資料探勘、服務水準

A Study for Analysis L/S of Freeway by Using Data Mining Techniques

Song-Wei Chu¹ You-Lin Lin²

Abstract

This paper will use continuous probe vehicle data mining method and technique of data mining, choice the basic road section between Taipei Yuanshan Interchange and Taoyuan Nankan Interchange of National Sun Yat Sen Freeway, constructs the estimating model of road service level for this study through a strict process of data selecting and data disposing. And, through strict test and verification, we enable this model to be applied practically on management and control of freeway; the result showed: Through a long-term probe vehicle data, we can not only effectively separate rush hour/off hour and service level under different time intervals and different date type in the research area calculating the service level of each space unit, but also find the change of velocity and the trend of rush hour/off hour are consistent with the results of previous researches. On the other hand, this study successfully proves that at the road section without VD, we can still estimate reliable average road section velocity and simultaneously, the calculated road service level can also be the reference for traffic management and administration by related units.

Keywords: Probe Vehicle, Data Mining, Level of Service

一、緒論

1.1 研究背景與動機

¹清雲科技大學行銷與流通管理系助理教授（聯絡地址:320 桃園縣中壢市健行路 229 號;電話:03-4581196 轉 7501;E-mail: swchu@cyu.edu.tw）

²清雲科技大學經營管理碩士（聯絡地址:320 桃園縣中壢市健行路 229 號;電話:03-4581196 轉 7501;E-mail: M9434010@cyu.edu.tw）

利用裝置具備全球衛星定位系統(Global Position System以下簡稱GPS)、無線通訊等功能車上單元(On Board Unit, 以下簡稱OBU)之探針車(Probe Vehicle)來蒐集交通資訊,隨著實際蒐集交通資訊探針車數量之增加以及相關理論與應用等研究成果之發展,而逐步變成即時交通資訊蒐集之主流。因此,若能透過探針車提供ATMS上所需之相關數據,實為成本較低且應用性更廣的方法。

回顧過去相關之文獻,發現以下幾點:1.傳統行駛時間調查方法之效率及準確度有待提升;2.車輛偵測器損壞率偏高,且佈設密度與數量不足,影響資訊蒐集精度:目前國內行駛速率資訊之計算依據,多以路面上佈設之車輛偵測器(Vehicle Detector, 以下簡稱VD)為主,但VD之埋設需經路面挖掘、VD佈設、重鋪柏油等程序,佈設不容易;且損壞率偏高,亦普遍存在維護困難的問題。3.智慧型運輸系統相關技術之演進,使探針車所能蒐集之資訊更為廣泛且即時。

1.2 研究目的

經由上述背景與動機之探討,本論文之研究目的在於利用現有裝置GPS之車輛做為探針車,再透過其所傳回之探針車資料,發展一套資料探勘的處理方式,藉由估算出之路段行駛速率,進行道路服務水準模式之構建,以做為道路服務水準評估資訊提供之用,使本研究之成果能夠真實地應用在現有道路路段服務水準的調查與評估,如此不但可省去傳統人工調查之工作與經費,亦可逐步取代現行推估路段行駛速率與道路服務水準的方法。

二、文獻探討

本研究旨在利用資料探勘技術,建構道路服務水準的評估方法。因此,根據本研究的主題,將先分別探討資料探勘之涵義、探針車蒐集交通資訊相關文獻與道路服務水準等文獻資料,藉此了解與研究主題相關的文獻與研究,以便在進行本研究時可增加對此一主題的解析。

2.1 資料探勘之涵義

Frawley(1991)對於認為資料探勘在資料庫中發掘出非顯然的、前所未有的及潛在的可能有用資訊之過程。Fayyad (1996)則認為資料探勘則是通過資料庫知識探索過程中特定的演算處理,在可接受的計算效率下生成特定模式的一個步驟。

綜合各學者給予的定義,「資料探勘」簡單來說:即是從資料庫中,利用各種分析法與技術,將過去所累積的大量歷史資料,挖掘出隱藏在大量資料中先前所不知道的和有用的資訊與知識,進行分析、歸納、預測及整合,以萃取出有用的資訊或找出有趣的樣式與知識,提供使用者或企業管理階層利用此資訊或知識作為訂定決策與問題解決的依據。

2.2 探針車蒐集交通資訊相關文獻

Sen (1997)主要探討探針車的數量(探針車經過多少路段)對於路段旅行時間預估的影響。Bruce (1999)分析討論樣本偏差對探針車輛所得之估計值準確性造成的

影響。運用車流理論來證明，抵達時間的分配所產生之偏差與各路段中探針車輛到路口轉彎所佔有的比例，會影響樣本估計值的平均數，並造成延誤。Frederick (2001) 展示一個新的演算法，運用大眾運具當作探針車輛，來判斷經過一條公路或是動態走廊所需的旅行時間與速率。Hellinga (2002) 將車輛旅行時間分為行駛時間及停等時間，在一號誌化的幹道上，假設探針車與整體車流平均旅行時間相同，所以只須考慮探針車與整體車流平均停等時間的差異。

2.3 道路之服務水準

依據交通部運輸研究所之定義(民國90)，服務水準指交通設施服務品質好壞之程度。目前容量分析之方法通常將服務水準分成數等級。最常見的分級包括A、B、C、D、E及F等六級。A級代表最佳之服務水準，F級則代表最差的服務水準。服務水準之劃分須根據績效指標。績效指標乃反應服務品質之交通參數。因不同交通設施有不同的作業性質，所以有必要利用不同的績效指標以劃分不同設施的服務水準。例如密度及平均速率常用於劃分高速公路之服務水準，而車輛之延滯時間則常用於劃分號誌化路口及收費站之服務水準。如下表一：

表一 高速公路服務水準等級之劃分標準表

服務水準	密度，D (小客車/公里/車道)	平均速率 (公里/小時)
A	$0 \leq D < 14$	≥ 90
B	$14 \leq D < 18$	≥ 85
C	$18 \leq D < 23$	≥ 80
D	$23 \leq D < 29$	≥ 70
E	$29 \leq D < 35$	≥ 60
F	$D \geq 35$	或 < 60

資料來源：交通部運輸研究所

三、研究方法

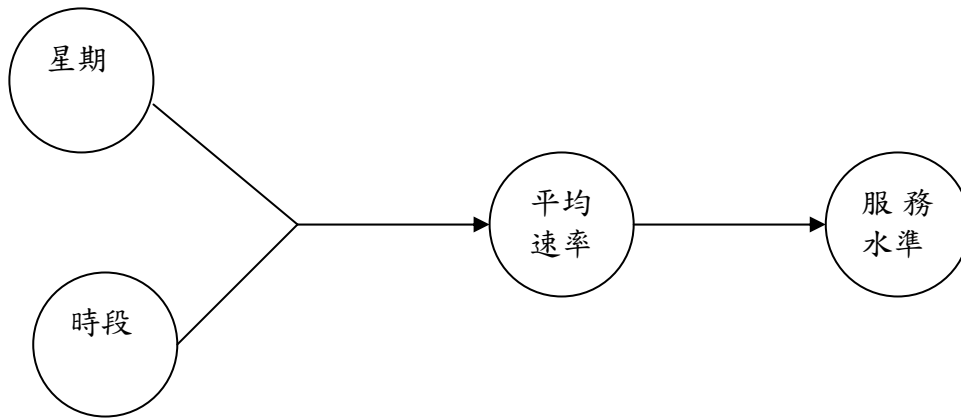
本研究主要針對前述研究目的與相關文獻的探討，提出本研究之研究方法。

3.1 研究對象與範圍

本研究以道路服務水準之評估為主，需取得道路密度（小客車/公里/車道）與平均速率等資料，且為兼顧本研究取得之探偵車歷史資料與車輛偵測器之密度等因素，故研究範圍將以國道一號台北圓山交流道至桃園南崁交流道之路段為主，且道路服務水準之評估，將以路段行駛速率為主要依據，而不針對道路容量作討論。

3.2 觀念性架構

根據文獻回顧之理論，擬定研究架構，從「星期」和「時段」二項獨立變數對「平均速率」之影響作為探討。第二部份則根據架構提出研究假設，以作為實證之依據。而運用資料探勘方法之中的多元迴歸分析，來建構道路服務水準模式、檢定且驗證其解釋結果。



圖一 本研究架構

3.3 資料分析方法

本研究之分析方法，依本研究的目的及假設之驗證需要，分成下列兩個階段：

- (一) 統計分析方法：利用 SPSS 10.0 版與 Excel 等工具軟體，進行路段行車速率計算、時間向度切割處理作業，且對道路服務水準模式進行比較驗證與檢定。
- (二) 資料探勘方法：資料經由相關之處理程序後，即可針對研究目的選取適用之資料探勘方法，資料探勘可使用之技術或方法相當多，如線性迴歸、邏輯斯迴歸、類神經網路、決策樹、群集分析或是關聯規則等，本研究以多元迴歸方法構建高速公路道路服務水準之模式，進而依據迴歸模式所輸入之變數，和參照道路服務水準評估準則，利用統計分析方法計算出之路段行駛速率及道路服務水準評估分析，並提出相關管理對策。

四、道路服務水準模式關係之建構

經由前述章節資料庫建立與處理確定後，即可確立所蒐集探針車資料紀錄之來源，故可對於所建立之探針車資料庫進行資料採礦，於本章節中，首先驗證依據該資料庫計算出之路段行駛速率，是否正確可信，以期能夠驗證說明路段行駛速率之準確度，爾後再以資料採礦技術中之多元迴歸分析法，對於探針車所紀錄的資料進行道路服務水準之模式構建與檢定驗證，進而確認道路服務水準變數之間的影響項目與關係。

4.1 道路服務水準路段行駛速率資料之驗證

為驗證依據該資料計算出之路段行駛速率，是具有準確度，本研究另選取在亞通客運上裝有數位式行車紀錄器之資料作為驗證比較。運用SPSS10.0版軟體之獨立樣本t檢定，對一天資料進行之驗證，其虛無及對立假設為，如下：

$$H_0: \mu_0 = \mu_1$$

$$H_1: \mu_0 \neq \mu_1$$

在 $\alpha=0.05$ 的顯著水準下，結果顯示 $F=0.270 < F(1-0.05)$ ，且顯著性(雙尾)為 $0.322 > \alpha=0.05$ ，故不顯著(如下表)，因此模式接受 $H_0: \mu_0 = \mu_1$ 之假設，表示本研究之探針車資料與亞通客運數位式行車紀錄器之資料是沒有差異的，且具有一致性。

表一 獨立樣本 t 檢定

	變異數相等的 Levene 檢定		平均數相等的 t 檢定		
	F 檢定	顯著性	t	自由度	顯著性 (雙尾)
假設變異數相等	0.270	0.580	0.982	1659	0.323
不假設變異數相等			0.991	1425	0.322

4.2 道路服務水準關係模式之設立

將設定模式中之變數為自變數或依變數，基於此部分所探討之目的，本研究將路段平均行駛速率作為依變數(Y)，將星期、時段作為自變數(X)，並分別構建多元迴歸式。迴歸模式架構如下：

$$\text{模式： } Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_{14}X_{14} + \varepsilon$$

4.3 道路服務水準關係模式之檢定

本研究運用SPSS10.0版軟體來建立迴歸模式，迴歸模式中選擇變項的方法有強迫進入法、向前選取法、向後選取法以及逐步迴歸法，由於本研究模式重點在於釐清研究變數間關係，以及如何對於依變數的變異提出一套具有最合理解釋的迴歸模式，且藉由迴歸模式之建立，得以發展出一套對於依變項的預測系統。所以先進行強迫進入法與逐步迴歸法，並觀察模式中：F值(F-test)判定是否具有統計意義，以及各項變數係數之t值是否顯著，以構建與檢定迴歸模式。F檢定值係檢定迴歸方程式，所有之係數是否有穩定性，亦即對迴歸方程式之適合度適切性大小；一般採用F分配檢定之，其自由度為 $n-p$ (n ：樣本數， p ：迴歸變數個數)，說明如下：

$$H_0: \text{迴歸係數等於零}$$

$$H_1: \text{迴歸係數不全為零}$$

在 $\alpha=0.025$ 的顯著水準下，模式中之F值檢定結果 $F=81 > F_{0.025}$ ，因此模式拒絕 H_0 之假設，故表示該模式皆具有統計上之意義。

表二 模式之變異數分析

模式	平方和	自由度	F 檢定	顯著性
迴歸	260837	11	81	0.000*
殘差	823308	2834		
總和	1084145	2845		

迴歸係數檢定之目的在檢定各個迴歸係數是否為零，若有顯著性，則表示該係數及其對應之解釋變數，可以解釋其對因變數之影響，一般採用t分配檢定之，說明如下：在 $\alpha=0.025$ 的顯著水準下，模式中之t值檢定結果，全部都顯著，且t值都大於零，因此模式拒絕 H_0 之假設，故表示該模式皆具有統計上之意義。

表三 模式之係數分析

模式	未標準化係數		標準化係數	t	顯著性
	B 之估計值	標準誤	Beta 分配		
常數	65.538	1.022		64.281	0.000*
星期一	-13.356	1.211	-0.258	11054	0.000*
星期二	-2.828	1.139	0.061	2.465	0.014*
星期三	-5.292	1.283	-0.092	4.118	0.000*
星期四	5.759	1.314	0.096	4.377	0.000*
星期六	3.350	1.355	0.053	2.465	0.014*
星期日	7.493	1.302	0.127	5.747	0.000*
時段一	10.689	2.958	0.060	3.611	0.000*
時段二	-9.310	0.802	-0.217	-11.637	0.000*
時段四	3.292	0.900	0.067	3.656	0.000*
時段六	4.895	1.321	0.064	3.700	0.000*
時段七	10.918	1.381	0.136	7.890	0.000*

4.4 道路服務水準關係模式之驗證

為驗證本研究之模式推估出路段平均行駛速率，由於前人相關研究均缺乏進行此類驗證，故選取北上資料庫2008/07/06~2008/07/12一星期所計算出路段平均行駛速率之資料來做驗證比較。運用SPSS10.0版軟體之中獨立樣本t檢定，對北上資料庫2008/07/06~2008/07/12一星期資料進行之驗證，其虛無及對立假設為，如下：

$$H_0: \mu_0 = \mu_1$$

$$H_1: \mu_0 \neq \mu_1$$

在 $\alpha=0.05$ 的顯著水準下，結果顯示 $F=0.951 < F(1-0.05)$ ，且顯著性(雙尾)為 $0.397 > \alpha=0.05$ ，故不顯著(如下表四)，因此模式接受 $H_0: \mu_0 = \mu_1$ 之假設，表示本研究之模式推估出路段行駛速率之資料與北上資料庫之一星期資料是沒有差異的，具有一致性。

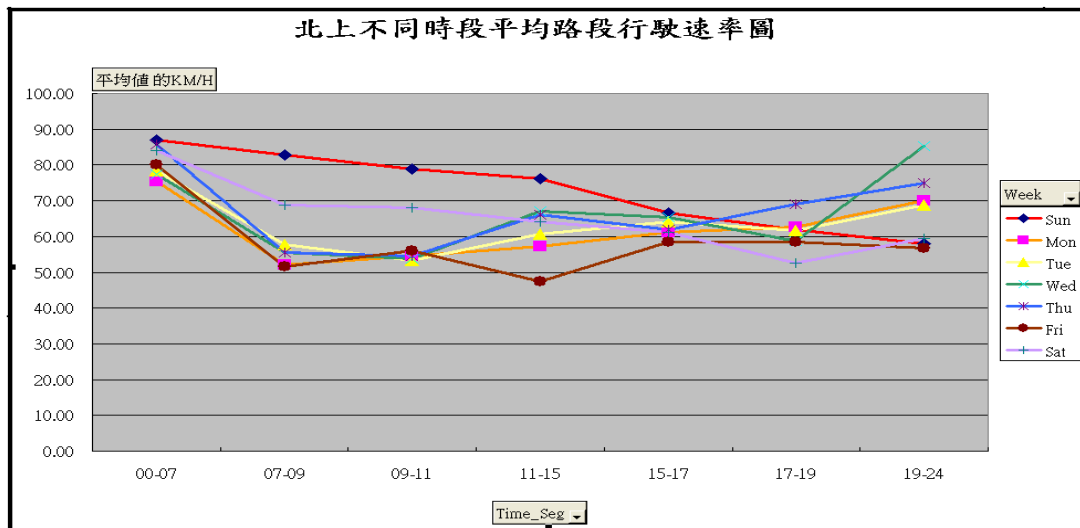
表四 獨立樣本 t 檢定

	變異數相等的 Levene 檢定		平均數相等的 t 檢定		
	F 檢定	顯著性	t	自由度	顯著性 (雙尾)
假設變異數相等	0.951	0.331	-0.850	96	0.397
不假設變異數相等			-0.850	94.231	0.397

本研究推估出之道路服務水準模式，其F檢定值為81，達到95%之檢定水準，而迴歸係數之t檢定也都達到95%之顯著水準，且模式驗證結果也與本研究資料庫一致性，所以代表本研究使用多元迴歸模式進行實證估計是適當的，推估出之道路服務水準模式是有說服力。

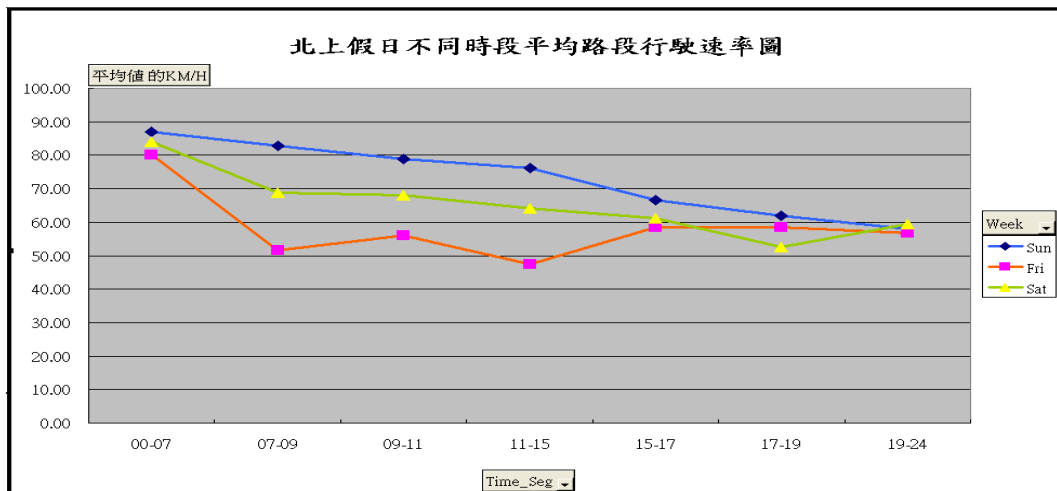
五、道路服務水準評估分析

將探針車資料經過資料清理與過濾後，分成北上路段資料庫。進一步將上述空間採樣點，改算成路段平均速率，由於各小時資料較為不足，且無法有效反映出尖峰/離峰路段速率之差別，故進一步將各小時依照交通部運輸研究所（2005）與趙志民（2004）在時段上歸類的方法，將資料聚累如圖二所示：

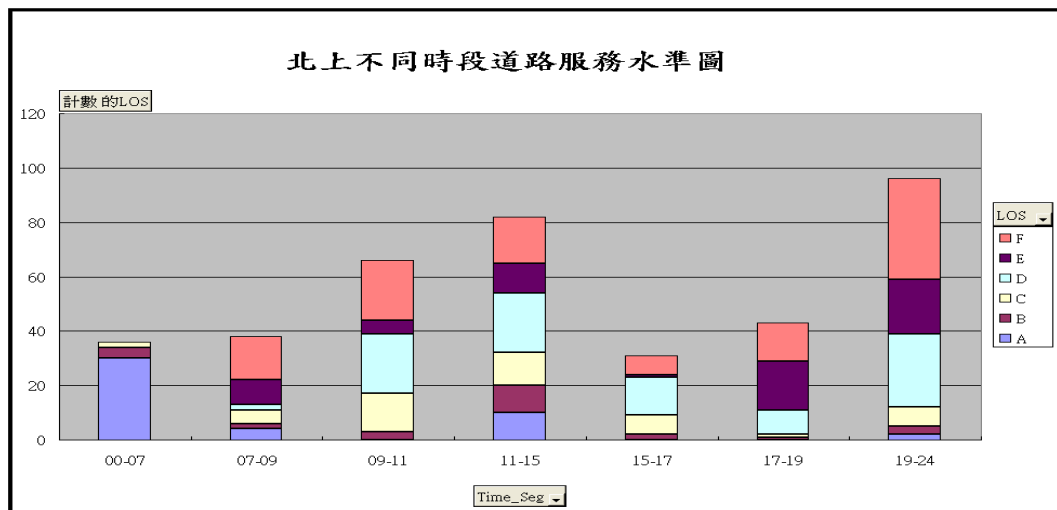


圖二 北上不同時段平均路段行駛速率圖

由圖 6-1 即可以明顯發現在時間上的尖峰與離峰對平均速率的影響極大，且本路段之尖峰出現在禮拜天 19-24；禮拜一 07-11、15-17；禮拜二 07-11、17-19；禮拜三 07-11、17-19；禮拜四 11-17；禮拜五 07-09、11-24 以及禮拜六 15-24，當中可以清楚看出，週五之平均路段行駛速率乃每週當中平均速率最低者，而週六、週日之平均速率則又有上升之趨勢。亦即每一週的尖峰時間均有所不同。進一步分成平日與假日(平日係指一般上班日，以行政院人事行政局公告為準，扣除鄰接假日之日程；假日則包括禮拜六、禮拜日以及國定假日)，來做比較，可以看到北上路段之禮拜六、禮拜日隨著假日結束，平均路段行駛速率也跟著降低，而禮拜五下班之返鄉旅次較多，故整天平均路段行駛速率普遍偏低，如下圖三：



圖三 北上假日不同時段平均路段行駛速率圖



圖四 北上不同時段道路服務水準圖

我們若進一步將平均時段速率依照標準轉換成服務水準，依不同時段分析之結果如上圖四所示：我們可以發現跟上述趨勢一致，且服務水準最高的時段都出現在離峰的 00-09 以及 11-15 時段，相反的尖峰時段的服務水準普遍較低。另一方面，若依各週不同日期類型來討論服務水準。

六、結論與建議

綜合前述實證分析之結果，具體提出本研究之結論，最後則提出未來研究建議，以供後續研究者參考。

6.1 結論

研究結果顯示透過較高時、空解像力之資料，不僅可有效區隔研究區內不同時段、不同日期類型的尖峰/離峰時間以及服務水準，而本研究之道路服務水準推估模式，經由嚴謹的檢定與驗證，可應用於高速公路之管理與控制上，在模式操

作過程之中，可以透過更新每月之探針車蒐集資料，套用在本模式，可換算出另一參數值，進而利用服務水準之分級標準，來判斷出服務水準等級，值得交通管理相關單位加以運用並參考。另一方面，本研究成功證明於VD缺乏之路段，亦可估算出可靠之平均路段速率，而本研究雖平均路段速率與前人研究相比，其平均速率較低，主要原因在探針車資料蒐集具有採樣數量與速率成反比之特性上，導致平均速率較低，但在速率變化以及尖峰與離峰之趨勢與前人研究成果一致。但本研究無論在時間或空間尺度上的研究成果，均較前人精細。

6.2 後續研究建議

- (一) 資料極值或錯誤之消除：初步研究成果發現部分極值或錯誤造成服務水準評估有所偏差，後續將與其他資料比對，剔除不正常數據。
- (二) 投入更多時空因子：利用空間關聯法則，找出尖峰／離峰、時段、日期類型、節慶、事故、大雨、坡度、彎道等對服務水準之影響模式，探討影響道路服務水準之顯性與隱性因子，並利用空間資料探勘所取得之知識，修正現行道路服務水準影響因子。
- (三) 增加道路服務水準推估模式之比較：如類神經網路、時間序列，進而比較道路服務水準推估模式。

參考文獻

1. Ashish Sen, Piyushimita Thakuria, Xia-Quon Zhu, Alan Karr, “Frequency of Probe Reports and Variance of Travel Time Estimates” , *Journal of Transportation Engineering*, Vol. 123, No. 4, pp. 290-297, July/August 1997.
2. Bruce R. Hellinga and Liping Fu, “Assessing Expected Accuracy of Probe Vehicle Travel Time Reports” , *Journal of Transportation Engineering*, Vol. 125, No. 6, pp. 524-530, November/December 1999.
3. Fayyad, U., G. Piatetsky-Shapiro and P. Smyth, “The KDD Process for Extracting Useful Knowledge from Volumes of Data” , *Communication of the ACM*, Vol. 39, No. 11, pp. 27 - 34, 1996.
4. Frawley, W. J., Paitetsky-Shapiro, G. and Matheus, C. J., “Knowledge Discovery in Databases: An Overview, Knowledge Discovery in Databases” , California, AAAI/MIT Press, pp.1-30,1991.
5. Frederick W. Cathey and Daniel J. Dailey, “Transit Vehicles as Traffic Probe Sensors” , *Transportation Research Record* 1804, pp. 23-30, 2002.
6. 交通部運輸研究所，2000年，台灣地區智慧型運輸系統綱要計劃，台北市。
7. 交通部運輸研究所，2001年，公路行駛時間調查(九十年)，台北市。
8. 交通部運輸研究所，2001年，台灣地區公路容量手冊，台北市。
9. 吳佳峰，2001年，有GPS資訊提供下之車輛旅行時間預估模式之研究，交通

大學運輸工程與管理學系碩士論文。

10. 陳佳麟，2003 年，*資料探勘於顧客關係管理應用之研究*，台北大學企業管理學系碩士論文。
11. 周文賢，2004 年，*多變量統計分析—SAS/STAT 使用方法*，台北市，智勝文化。
12. 趙志民，2004 年，*高速公路路段速率推估之研究*，交通大學管理學院碩士在職專班運輸物流組碩士論文。