

行政院國家科學委員會專題研究計畫 成果報告

子計畫三：Telematics 服務系統之建構與應用(I)

計畫類別：整合型計畫

計畫編號：NSC93-2218-E-009-049-

執行期間：93年08月01日至94年07月31日

執行單位：國立交通大學經營管理研究所

計畫主持人：毛治國

計畫參與人員：邱孟佑

報告類型：精簡報告

處理方式：本計畫可公開查詢

中 華 民 國 94 年 9 月 30 日

摘要

異質資訊(Heterogeneous Information)融合與應用之核心技術平台的開發，將可協助先進用路人資訊系統(Advanced Traveler Information System, ATIS)之具體發展，符合用路人對各項旅運資訊的需求，並將影響用路人旅運決策行為。國內雖有部分研究計畫探討用路人資訊需求項目、ATIS 功能規劃與產品類型，惟較少研究以宏觀系統服務的觀點，完整規劃用路人異質資訊功能模組，並實際開發建置整合性核心技術平台。

爰此，本整合型計畫之「研究總目標」在於透過用路人異質資訊需求評估、資料融合、資料探勘、資訊分析、策略研擬、系統規劃與雛型開發等過程，最終目標期能建置及展示一套整合性之用路人異質資訊應用平台，滿足用路人對各項異質資訊之需求，進而達到提升運輸效率與安全的目的。

關鍵詞：異質資訊；資訊融合；先進用路人資訊系統

Abstract

This study is to develop the technical platform for data fusion and applications of heterogeneous information, then assist the implementation of ATIS. In order to develop and display an integrated platform for traveler heterogeneous information, it will be accomplish by a series of process, such as demand evaluation, data fusion, data mining, information analysis, strategy planning and prototype development. This platform can meet the demand of travel information, and affect the traveler behavior.

Keywords : Heterogeneous Information, Data Fusion, Advanced Traveler Information System (ATIS)

一、前言

根據交通部於 93 年 7 月的統計，自用小客車及營業用小客車總數共有 5,316,469 輛，截至 94 年 7 月止，其自用小客車及營業用小客車之總數量已達 5,566,585，其較上年同月評比成長 4.7%，該數據顯示汽車數量不僅持續成長，其背後更衍生出車輛成長帶來更多的使用者需求及交通壅塞等問題。雖然在國內已有多位學者在資訊科學的領域對如何發展一資訊提供平台與資料探勘技術已有所探討，但是鮮少針對 Telematics 相關議題進行研究，目前在 Telematics 上之研究多將車輛視為運輸工具，以駕駛人對車輛上的觀點進行相關議題之討論，其中以如何改善交通運輸狀況及行車便利性為主要探討議題，然而此類研究只能改善行車中浪費時間在等待到達目的地之問題上，無法提供用路人一有效利用時間於行車過程之解決方案，因此本計劃以用路人之需求為出發點，並針對交通運輸的需求來加以討論，期提供用路人在行車過程中所需資訊，以豐富駕駛人之行車生活。

二、計畫目的

本計劃目的為提供用路人一豐富及舒適之乘車生活為主旨，將第一、二子計畫所提供的資訊及其他相關資訊加以融合，建立一完整資訊提供的整合平台(Telematics Center)。在本計劃中我們利用資料探勘的技術，針對駕駛人及車內乘客所需進行需求分析，以期有效降低行車時浪費時間於等待上及提升用路人之行車資訊的切確性。因此，本計劃針對較急切的需求進行研究，以最佳路徑分析、停車場停車位預測及客製化分析、廣告行銷等議題作為第一年研究之計劃，期以本計劃之研究探討之研究成果，提供一相關技術移轉給國內的加值業者外並帶動國內相關產業的發展。

三、文獻探討

資料探勘(Data Mining)指由大量資料中擷取出有出價值之知識，亦即將資料轉換成知識的行為過程。這些資料包括一般食衣住行育樂資料、交通資料或多媒體資料，而知識則是資料間隱含關係的具體表達與呈現。而“資料間隱含關係”的不同，學者陳彥良[1]提出了八種不同的資料關係類別，分別是資料關聯性、順序性、結構性、週期性、類似性、有趣性、個人性、合用性等，其中，就資料關聯性之研究即挖掘關聯規則(Association Rule)、分群技術及分類技術為最重要的資料探勘技術。本計劃除利用資料探勘之相關技術外，也同時利用類神經之發現法則以導出其預測模型。

(一) 關連分析

資料關聯性之研究即在獲取資料間的關聯規則(Association Rule)，為最重要的資料挖掘問題之一。它的目的是要從眾多資料中，發現項目(Item)間的關聯。例如若在許多記錄中，我們發現某些項目的出現會引發其他項目的出現，這樣的關聯關係即可以用關聯規則的型式加以表達。例如：牛奶 \Rightarrow 麵包。

在探討關聯規則的挖掘之前，我們必須先了解最小 Support 和最小 Confidence 的概念，最小 Support 界定一個規則必須涵蓋的最少資料數目，最小的 Confidence 則界定這個規則

的預測強度。規則的 Support 和 Confidence 可以評估規則是否有趣，當挖掘演算法所找出的規則滿足使用者訂定的最小 Support 和 Confidence 的門檻時，這個規則才算成立。

(二) 分群

資料分群(Cluster)的主要精神就是根據資料分群的原則，將資料庫中的資料分類成群。所謂資料叢集的原則，即為“群組內的資料相似度最高，群組間的資料相似度最低”[4]。有許多資料叢集的演算法被提出，J.B.MacQueen[4]提出的 K-Means Method 是利用具代表性的樣本點將資料先予以分群，再找出各群資料的質心、群組界線，一步步地修正資料所隸屬的群組，直到精確分群為止。

(三) 分類

資料分類 (Data Classification) 是根據事先建立的分類模型或法則，將資料庫中的資料依其性質加以分門別類的過程。我們可分析現有的歷史資料或是將一完整資料庫做部分取樣，以找出資料的特徵，再依這些特徵對其它未經分類的資料作預測。資料分類技術常常被用來處理郵寄對象篩選的問題。J.R.Quinlan 提出了一個建構分類模型的方法—ID3[2]。它是以決策樹來表示分類規則的遞迴結構，使用由上而下 (Top-Down) 和各別擊破 (Divide-and-Conquer) 的策略，並使用資訊理論 (Information Theoretic) 的方法，建立出有效率的分類模型，將資料物件隨著樹的成長，逐步分割成更小的子集合，以達到分類的目的。

(四) 尋找相關性

尋找相關性演算法(Find Dependencies)主要是用來快速尋找資料的關聯性和相關性而不是這些相關性中找出其特性。換句話說，該演算法主要在找尋屬性與屬性間的關聯性強烈，以獲得相相關性高的子集合，因此，在資料探勘中使用尋找相關性演算法可以排除一些偏離值。ARNAVAC[5]為尋找相關性演算法中重要方法。

四、研究方法

如圖 1 所示，Telematics Center 內融合了交通部運研所台灣地區之路道地圖資料、即時交通資訊、台北市停車管理處之停車場資訊、子計劃一二中代理人所匯集之食衣住行育樂資料以及各駕駛之 Log 記錄，這些用路人所需之資料均被整理融合儲存於關聯式資料庫內 (RDB)，資料分析伺服器將藉由物件導向連結(OLE DB)中介軟體(Middleware)連結資料庫內資料，最後再由線上即時分析(OLAP)資料探勘模組將分析過濾比對後之結果與資料探勘應用子系統相結合而構成一資料探勘應用系統。本計劃利用 Telematics Center 所採集資料進行分析，以 PolyAnalyst[3]進行分析，其分析主題包含：最佳路徑分析、停車場停車位預測及客製化分析、廣告行銷等議題。在本計劃中，我們利用 ARNAVAC 演算法以求出各屬性質之間關係之強度並移除其偏離值，其後將這些顯示關連性之屬性進行探勘動作，由於各分析之主題不同，我們在各主題中將使用不同之分析方法以求得資料之間的關連性，再利用 Find Laws 演算法[3]以求得其預測模型，最後我們再針對各預測模型進行檢測動作，以篩選有價值之模型供路用人使用。其主要分析流程如圖二所示。

圖一 資料探勘應用處理流程

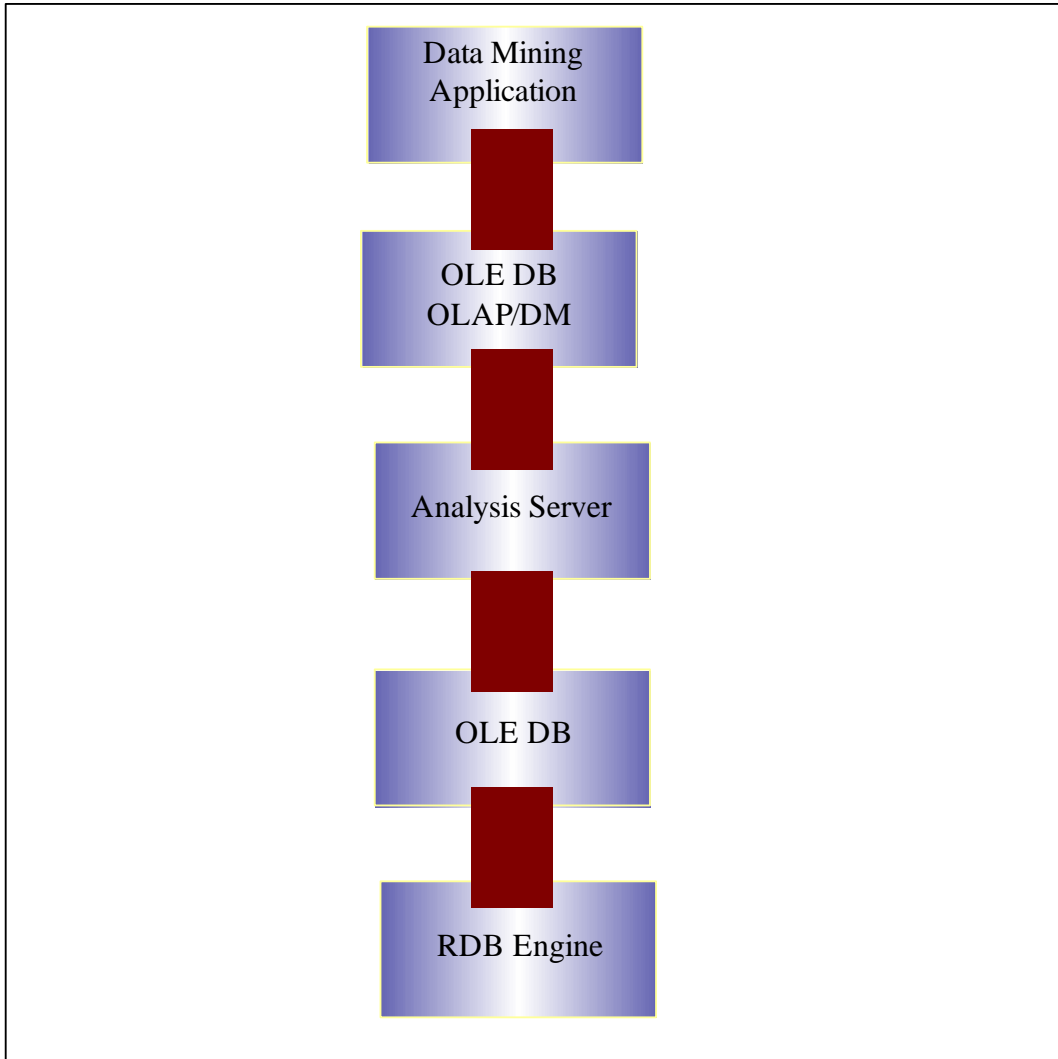
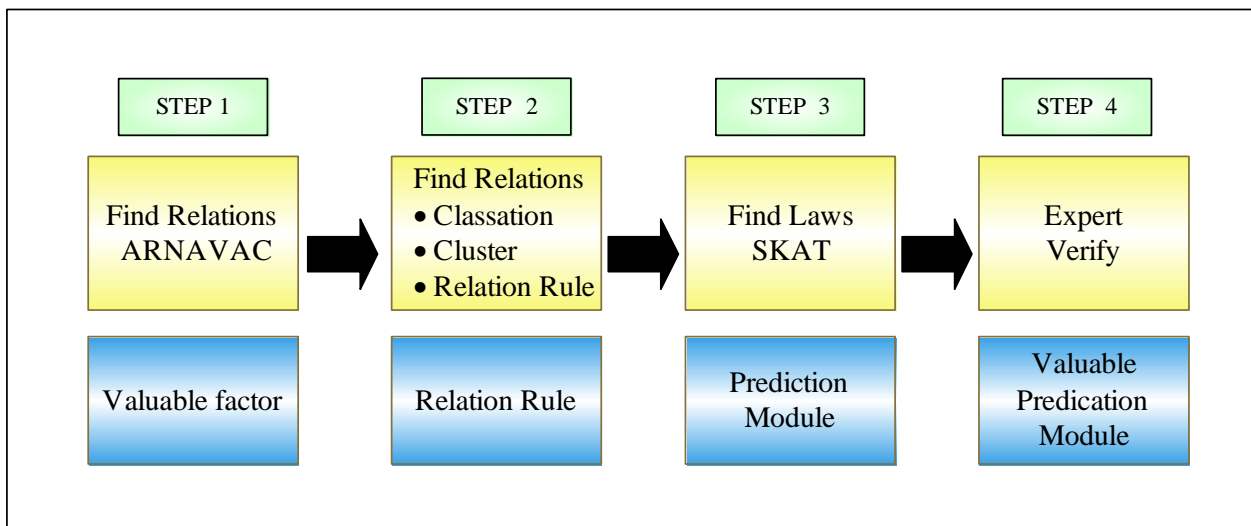


圖 2 資料探勘處理流程



(一) 最佳路徑預測：

目前大部份的研究皆以最短路徑為主要探討主題，惟最短路徑只計算兩點之間的最短距離欠缺即時路況資訊，故最短路徑不足以代表其為最佳路徑，因此本計劃中以為各路網結構為基礎，將 Telematics Center 中的道路資料之路結構搭配即時路況資料及各車機目前所在位置進行分析，先應用 PolyAnalyst 之尋找相關性演算法(Find Dependencies)以取得 Telematics Center 中各屬性與行車速度之間的相關性，藉以去除非影響行車速度之偏離屬性，再者，由 Telematics Center 中，擷取出影響行車速度之屬性資料，利用 Find Laws 演算法求出其各路段之行車時間預測模型，以作為未來車機之行車路徑之時間預測。在最佳路徑預測上，用路人需先定義其目的地地址，本系統在實在作上，利用 ArcView 建構出各道路之路網結構以計算出其可行路徑，其後，再利用預測模型以計算出各可行路路徑之行車時間以求得最佳路徑，再建議使用者行駛。

(二) 停車位預測：

停車位預測需搭配合前一功能：行車時間預測模型以進行預測，首先車機需利用 GPS 找出目前所在位置，其後選擇欲查詢之停車場或目前該車機行駛附近之停車場資訊，在本研究中，將 Telematics Center 中台北市停車管理處之停車位管理系統之車量的進出記錄及目前停車位狀況及到達目的地之時間預測加以整合，先行利用行車時間預估模型以求得到達其停車場之所需時間，再利用尋找相關性演算法(Find Dependencies)，以取得在一時間序列中各時間點影響停車位之相關屬性。在取得影響停車位之因子後，再利用 Find Laws 建立停車場預估模型以幫使用者預測各停車場之停車狀況。

(三) 客製化分析：

客製化分析的目的旨在提供駕駛人一個人化之行車資訊，客製化的分析分為兩個階段，一為系統資訊之分類，一為用路人之行為分析。在系統資訊分類中，本計劃將 Telematics Center 中欲呈現給用路人之資分類為食衣住行育樂等五大類別，並在各類別中區分為更細之子類別，如表 1 所示。本研究利用 HMM 演算法，先將 Telematics Center 中各子類別建立一關鍵字庫，其後，再將擷取代理人所擷取到的資訊，利用關鍵字比對法進行比對，最後將該則資訊分類入關連性高之類別。在用路人之行為分析方面，用路人在使用車機系統的同時，系統會記錄其所流覽之網頁資訊及其經常行使之路徑，由於各網頁之資訊於融入 Telematics Center 之同時，已先行進行分類動作，故當使用者點選其網頁時，即以記錄該資訊所屬之類別，其後再利用分群演算法(Cluster)以分析各用路人之所屬類別。因此，在本分析上，我們根據使用者平時使用車機之行為記錄，分別計算出該用路人與各類別之間的關連性，再下一次用路人再使用相關資訊時，即優先提供與用路人平時使用習慣最高的類別新聞給予用路人，例如一喜歡郊外旅遊之用路人，因經常流覽旅遊類別之網站，因其活動之區域北部地區，故當用路人欲流覽網站時，則預先提供北區旅遊之新聞給與該用路人。

類別	大類別	子類別
食	料理	中式、西式、速食
	夜市	觀光夜市
	流行食尚	甜點、熱食
衣	流行、時尚	彩妝保養、服飾造型、明星、時裝
住	旅館	民宿、飯店
	酒店	
行	旅遊	台灣行腳、風景區、名勝古蹟
	農場、牧場	
育樂	表演	現代戲劇、傳統戲劇
	展覽	美術、博物
	休閒	溫泉、登山、攝影、釣魚、繪畫、音樂、電玩、電影、唱歌、遊樂園
	比賽	棒球、籃球、運動會
	消費	百貨、賣場、超市、
	新聞	政治、國際、兩岸、財經、影視、體育、生活、休閒、健康、社會
	科技	電腦、手機、數位生活

表 1：食衣住行育樂分類表

(四) 廣告行銷分析

廣告行銷如同客製化分析一樣，我們先將各則廣告類別根據食衣住行育樂等五大類別加以分類，由於用路人之行為特性已透過 Cluster 分群方法進行分類動作，故在廣告發送的上選擇對該廣告之所屬類別相關性高之用路人作為廣告發放對象，當用路人行經過該店家時，則可提供該用路人相關廣告訊息。由於目前車機接收訊息來的來源方式仍以 GPRS 為主，然而目前之 GPRS 之費率計算封包數為其計算方式，故利用本計劃所提出之分群方式，可有效減少廣告商發送廣告訊息所產生之廣告成本，並且也有有效替用路人抵擋垃圾廣告訊息的麻煩。因此利用廣告行銷分析能針對目前地點、時間、篩選並傳送適當的廣告，減少使用者對廣告的反感並提高廣告發送的成本效益。

五、結果與討論

本計畫利用已融合之各類資訊進行分析應用，首先產生路徑的旅行時間推估，並與先進電信資訊 Telematics 的服務整合，其應用層面將可擴及 ITS 各發展領域。由於未來電子收費系統的推廣普及，因應電子收費需求用路人必須裝設有 VPS(Vehicle Positioning System) 車機，透過車機之使用，我們將可提供一個人化之行車資訊給予用路人，並且將其所流覽食、衣、住、行、育樂等資料加以記錄，以整合為完整的電信資訊(Telematics)服務系統。

在本計劃之執行第一年，以先行針對路徑規劃建立一旅行時間預估模型，並實際應用

於車機中，如圖 3 所示。在用路人行駛過程中，僅只需點選出目的地地址，本系統即可藉由該模型及路網結構導出之可行性分析進行推估避開雍塞及意外事路路段，以推薦出一最佳路徑導引。在停車場停車位預測上，我們提供了一停車位預測模型，用路人僅需提供所欲停放之停車場位置，本系統即會先行預測到達該停車場之旅行時間，其後，在以到達該停車場之時間為主題下進行停車位之預測，其實作如圖 4 所示。至於客製化及市場分析，其主要針對各用路人之行為記錄及 Telematics Center 中之資訊加以分類，客製化之目的旨在提供各用路人之個別化之 VPS 操作環境，以其需求為出發點，優先提供該用路人有興趣之議題，其成果如圖 5 所示。最後，廣告分析則為客制化之延伸應用，如圖 6 所示，其結合了客制化分析時各用路人與各分類之間的關係強度加以分類，並將廣告商欲發佈之廣告進行分類動作，在駕駛人行駛於各路段時提供其有興趣之廣告議題，並減少廣告商之廣告發送成本。

因此在本計劃中，主要目標在於首先融合異質交通運輸資訊成為路段的旅行時間，其次結合日趨成熟之 VPS 車機與 Telematics 服務平台之規劃研究來提出一個別式用路人路徑導引機制，用路人可以藉由 VPS 車機得到正確的道路交通資訊及其個人化之需求資訊，並選擇最適合自己的替代方案作最佳的反應控制策略，藉以提昇運輸效率。

圖 3 最佳路徑之導引圖



圖 4 停車位定位及預測畫面



圖 5 個人化之新聞資訊圖



圖 6 目標行銷之廣告服務圖



六、參考文獻

- [1] 沈清正、陳仕昇、高鴻斌、張元哲、陳家仁、黃琮盛、陳彥良。資料間隱含關係的挖掘與展望。資訊管理學報，9（專），75-100,2002。
- [2] J.R Quinlan, "Induction of Decision Tree", Machine Learning, Vol.1, pp.81-106,1986
- [3] Megaputer. Polyanalyst 4.6, 2005. <http://www.megaputer.com/>
- [4] Ming-Syan Chen, Jiawei Han, and Philip S. Yu, "Data mining: An Overview from a Database Perspective", IEEE Transactions on Knowledge and Data Engineering, Vol.8, No.6, December 1996
- [5] P. Mitra, G. Wiederhold, and M. Kersten. A graph-oriented model for articulation of ontology interdependencies. In *Proceedings of Conference on Extending Database Technology (EDBT 2000)*, Konstanz, Germany, March 2000.