行政院國家科學委員會專題研究計畫 成果報告

以資料探勘技術辨識混合車流中駕駛行為類別之研究
研究成果報告(精簡版)

計 畫 類 別 ：個別型
計 畫 編 號 ：NSC 96-2415-H-216-001-
執 行 期 間 ：96 年 01 月 01 日至 96 年 10 月 31 日
執 行 單 位 ：中華大學運輸科技與物流管理學系

計 畫 主 持 人：羅仕京

計畫參與人員： 大學生-兼任助理：王曉惠、蔡筱葳、陳孟曦

報 告 附 件 ：出席國際會議研究心得報告及發表論文

處 理 方 式 ：本計畫可公開查詢

中 華 民 國 96 年 10 月 30 日

行政院國家科學委員會補助專題研究計畫 ■ 成 果 報 告
□期中進度報告

以資料探勘技術辨識混合車流中駕駛行為類別之研究

計畫類別：■ 個別型計畫　　□ 整合型計畫
計畫編號：NSC 96－2415－H－216－001－
執行期間：民國 96 年 1 月 1 日至　96 年 10 月 31 日

計畫主持人：羅仕京
共同主持人：
計畫參與人員：蔡筱葳、王曉惠、陳孟曦

成果報告類型(依經費核定清單規定繳交)：■精簡報告　□完整報告

本成果報告包括以下應繳交之附件：
□赴國外出差或研習心得報告一份
□赴大陸地區出差或研習心得報告一份
■出席國際學術會議心得報告及發表之論文各一份
□國際合作研究計畫國外研究報告書一份

處理方式：除產學合作研究計畫、提升產業技術及人才培育研究計畫、列管計畫及下列情形者外，得立即公開查詢
□涉及專利或其他智慧財產權，□一年□二年後可公開查詢

執行單位：中華大學運輸科技與物流管理學系

中　華　民　國九十六年十月二十九日

一、中、英文摘要

中文摘要

道路駕駛行為為一複雜之研究課題，而要獲得準確之車流速度、密度與流量預測必須構建具有描述多種用路者之車流模式[Helbing, 2001, Hoogendoorn and Bovy, 2000, Lo, 2002]。然而，以往用路者行為通常以車種決定，如小汽車、大客車與貨車等，但把同一種車輛的使用行為視作一致，並不完全合適。因不同性別、年齡、旅次目的、載客人數與駕駛車輛廠牌等，都會影響用路行為。若以傳統問卷調查方式，調查對象、訪談者與模式構建、校估者可能有認知上的不同，而產生偏差。因此，本研究擬以資料探勘技術(data mining)由既有車流資料當中萃取不同駕駛行為之種類，進而校估描述不同用路行為之參數。本研究之基本假設為各種用路行為所組成之車流滿足高斯分配(即常態分配)[Helbing, 2001]。基於以上假設，本研究以期望值最大(expectation-maximization) 技術，訓練學習並辨識不同駕駛行為，並以兩演算例比較辨識結果。根據模擬結果，本研究所提出之方法成功地辨識出用路行為的種類。若將此方法與偵測器資料結合，不僅節省調查訪問成本，亦可進一步構建自動校估機制處理資料提供車流模式預測所需之參數。

關鍵詞：資料探勘、期望最大化、高斯混合、用路行為、車流模式。

英文摘要

Understanding driving behavior is a complicated researching topic. To describe accurate speed, flow and density of a multiclass users traffic flow, an adequate model is needed. [Helbing, 2001, Hoogendoorn and Bovy, 2000, Lo, 2002]。User’s classes are determined by types of vehicles in previous studies. However, considering all drivers with the same type of vehicles have the same behavior is too rough for traffic flow study. Conventionally, classifying driving behaviors is obtained by inquiring from door to door. It takes a lot of cost and may produce bias because of the different agreement among the inquirers, drivers and researchers. Therefore, a new method, which is based on data mining technique, is proposed to classify driving behavior in multiclass user traffic flow. In this study, driving behaviors are assumption to be in the form of Gaussian distribution [Helbing, 2001]. According to the assumption, expectation-maximization method is employed to train and classify different driving behaviors. By the method, a cost saving and automatic way for traffic data processing and parameter extracting is obtained.

Keyword：data mining, expectation-maximization, Gaussian mixture, multiclass user, traffic flow model.

二、報告內容

(一)、前言

運輸領域之研究在於使人或物以最方便、快速、經濟、安全與舒適的方式到達目的地，而在講求效率與競爭力的現代，提昇運輸效率不僅能降低營運成本與旅運者的時間成本，同時也能減少延滯所產生的外部成本。為提昇運輸系統效率，即時收集並預測交通資訊進而研擬交通控制策略為重要的方法之一，要獲得預測之交通資訊與控制方案評估，需藉由構建並求解動態交通模式達成，動態車流模式即為其中之一。而為獲得正確的預測，所構建之模式需能詳細地描述道路車流狀況與組成，因此有多種用路行為多車道之車流模式之發展。多種用路行為車流係指當路當中同時有多種駕駛行為的車流，所涵蓋的範圍較多車種廣泛，因為同一車種可能也有不同用路行為。一般而言，多種用路行為車流模式為車流模式之延伸，也就是基本車流模式相同，但各種不同用路行為用不同的參數描述，如何較估參數並決定各種用路行為的當量數(或權重)即為一重要的研究課題。

(二)、研究目的

以往進行若以傳統的問卷調查方式調查用路行為，不論是郵寄、電話或家戶訪問都需耗費大量的時間、金錢與人力成本，而且收集得來的結果可能會產生偏差，因為問卷設計

者、訪談者、受訪者、模式構建者與校估者之間的認知可能不同，造成訪問結果的差異與參數的偏差。再者影響用路行為的因素很多，舉凡：性別、年齡、工作、旅次目的、乘載人數、氣候、道路種類與型態、車輛廠牌等都會影響表現出來的用路行為，使問題更顯得複雜。在模式發展的過程中，根據經驗與理論推導，均衡時各用路行為之車流現象可用常態分布(高斯分布)描述。以此為基礎，整體車流現象可視為多個高斯分布的混合，若能將整體車流資料中辨識出當中含有幾個高斯分布與混合之權重，便能得到車流中有幾種用路行為及其參數與混合比例。本研究即以此假設，利用資料探勘(data mining)中之同質分組(clustering)技術與期望最大化法(expectation-maximization)，直接由車流資料中辨識獲得車流模式所需的參數。

(三)、文獻探討

3.1 多用路行為車流模式

車流理論為交通運輸領域中應用廣泛的研究課題，本研究之車流模式以巨觀模式為主，因為微觀模式所需之參數更為龐大。動態巨觀模式主要以 LWR 模式為基礎，LWR 模式是 Lighthill, Whitham [Lighthill and Whitham, 1955] 與 Richards [Richards, 1956] 由車輛數守恆推導出連續方程式（continuity equation）來描述車流行為，主要是用來描述車流密度隨時間變化的傳遞現象。此模式之基本假設有二：(1)車輛數守恆；(2)密度與速度之間存在一對一的函數關係。

LWR 模式如公式(1)所示：

$$\frac{\partial k}{\partial t}+\nabla\cdot Q=0 \qquad (1)$$

其中 k 為密度，Q 為流量。若密度與速度(u)呈線性關係，加上 $Q=ku$ 關係式，公式(1)可利用特徵線法(characteristic line method)與起始條件求出解析解。若密度與速度之關係式為非線性，則需藉助數值模擬求解。LWR 模式中所給定之密度速度關係式為靜態關係式，係指若密度變化，速度將立即隨之改變，與實際車流行進時會有反應時間延滯的情況不符。Payne [Payne, 1979] 引用一運動方程式(或稱為動量守恆式)取代原本之密度速度關係式，改善此一缺點，其模式如下：

$$\frac{\partial k}{\partial t}+\nabla\cdot Q=0\ , \qquad (2)$$

$$\frac{\partial u}{\partial t}+u(\nabla\cdot u)=-\frac{1}{k}\nabla\cdot\left(P_e(k)\right)+\frac{1}{\tau}\left(u_e(k)-u\right)\ , \qquad (3)$$

其中 $\bar{u}$ 為平均速度，τ 為反應時間，$u_e(k)$ 為均衡速度，$-\nabla\cdot\left(P_e(k)\right)/k$ 為預期反應項(anticipation term)，$P_e(k)$稱為均衡交通壓力(equilibrium traffic pressure)，受速度變異(speed variance)所影響。Michalopoulos 等人[Michalopoulos 1980, 1980, 1981, 1984, 1993]則利用黏滯項、半黏滯項與離散公式改善 Payne 之模式以增加計算速度。Helbing[Helbing, 1995, 1996, 1997, 2001] 進一步考慮隨時間變動之速度變異，導出氣體動力車流(gas-kinetic)模式：

$$\frac{\partial k}{\partial t}+\nabla\cdot Q=0 \qquad (4)$$

$$\frac{\partial u}{\partial t}+u(\nabla\cdot u)=-\frac{1}{k}\nabla\cdot(k\theta)+\frac{1}{\tau}\left(u_e(k)-u\right)+\frac{\mu}{k}\nabla\cdot(\nabla u) \qquad (5)$$

$$\frac{\partial \theta}{\partial t}+u(\nabla\cdot\theta)=-2\theta(\nabla\cdot u)+2\frac{\mu}{k}(\nabla\cdot u)^2+\frac{2}{\tau}\left(\theta_e(k)-\theta\right)+\frac{\mu}{k}\nabla\cdot(\nabla u)+\frac{\kappa}{k}\nabla\cdot(\nabla\theta) \qquad (6)$$

其中θ 為速度變異，μ與κ為參數。考慮了速度變異的影響，Helbing 的模式可描述：(1)車流密度高的區域，速度較低且速度變異較小；(2)車流密度低的區域，速度較高且速度變異較大；(3)速度變異的極大值發生在車隊後方速度最快的地方。

在 Helbing 的研究中發現，更高階的變數已不具顯著的影響，因此暫時沒有繼續發展高階車流模式的必要。由此可知，並非所有的車流情況均需應用 Helbing 之方程組，需視車流變數是否具有顯著影響而定。

若考慮多種用路行為模式則是將單一車流模式擴充成多組方程組，各種車流受各自的

運動方程控制同一道路中競爭前進[Hoogendoorn and Bovy, 1998, 1999, 2000]。以氣體動力模式為例：

$$\frac{\partial k_i}{\partial t}+\nabla\cdot Q_i=0 \tag{7}$$

$$\frac{\partial u_i}{\partial t}+u_i(\nabla\cdot u_i)=-\frac{1}{k_i}\nabla\cdot(k_i\theta_i)+\frac{1}{\tau_i}(u_{ie}(k_i)-u_i)+\frac{\mu_i}{k_i}\nabla\cdot(\nabla u_i) \tag{8}$$

$$\frac{\partial\theta_i}{\partial t}+u_i(\nabla\cdot\theta_i)=-2\theta_i(\nabla\cdot u_i)+2\frac{\mu_i}{k_i}(\nabla\cdot u_i)^2+\frac{2}{\tau_i}(\theta_{ie}(k_i)-\theta_i)+\frac{\mu_i}{k_i}\nabla\cdot(\nabla u_i)+\frac{\kappa_i}{k_i}\nabla\cdot(\nabla\theta_i) \tag{9}$$

其中下標 i 表示用路行為 i。若有 n 種行為，則有 n 組方程組。總密度表示成

$$k=\sum_{i=1}^{n}e_i k_i\ , \tag{10}$$

其中 e_i 表示第 i 種用路行為之當量數。在[Lo, 2002, Cho and Lo, 2002]中則考慮各車輛間之互動，以一擴散模式描述混合車流之現象，如公式(11)

$$div\mathbf{E}=-\Delta\phi=\sum_i\left[\frac{e_iK_{i0}}{\varepsilon_i}\exp\left(-\frac{e_i\phi-e_i\psi_i}{\Theta_{ie}}\right)\right]-\frac{e}{\varepsilon}k_s+K_a\ , \tag{11}$$

其中 $\mathbf{E}$ 為交通場(traffic field)。由以上回顧，不論何種混合車流模式，均需校估許多參數，並決定有多少種用路行為在車流當中。本研究以資料探勘方式直接從資料中萃取，不僅節省時間與金錢成本，亦可整合於車流模擬模式與程式之中。

3.2 資料探勘

資料探勘[Adriaans and Zantinge, 1996, Roiger and Geatz, 2003, Westphal and Blaxton, 1998, Fayyad, et. al, 1996, Freitas, 2002, Han and Kamber, 2001, Hand, et. al, 2001, Trueblood and Lovett, 2001]，為一種從所收集得來的資料中，分析並擷取有用資訊的技術，主要目的在於從雜亂無章的資料當中，藉著觀察趨勢與隱含的類別，了解原始資料中所透露的有用資訊，潛在的模型與有用的規則，而作進一步應用。目前的資料探勘技術均使用歸納法學習(induction-based learning)，利用既有資料，建構出資料當中的規則與訊息。一般而言，資料探勘先把原始資料分成訓練集合與測試集合。利用訓練集合學習分類資料；以測試集合測試模式正確性。簡言之，可將資料探勘歸納為下列四個步驟：(1)初步整理所收集的資料；(2)以資料探勘演算法學習並驗證；(3)解釋結果並分析；(4)應用所得資訊。資料探勘處理模式以圖 1 表示。

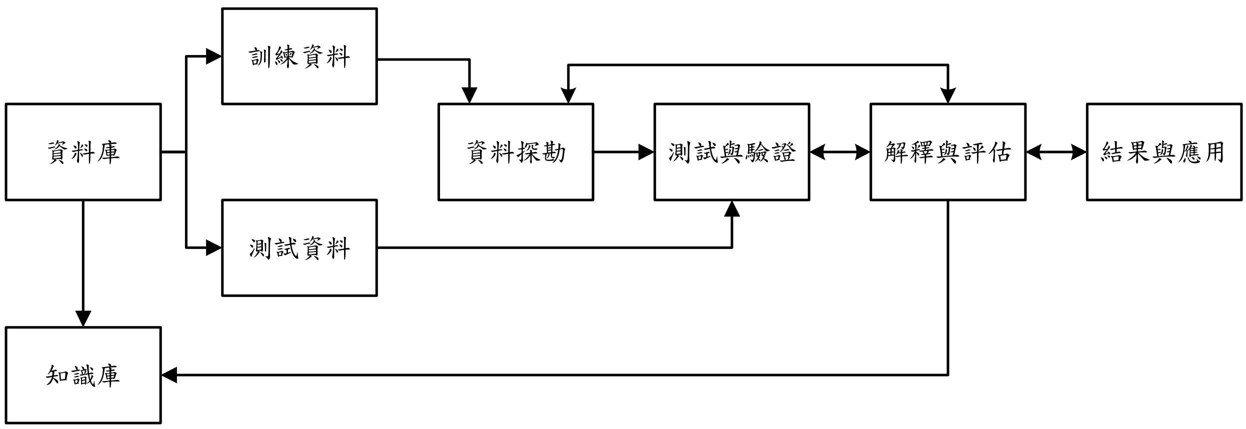


圖 1 資料探勘處理程序

資料探勘技術可概略分為監督式(supervised)與非監督式(unsupervised)兩種。監督式法主要是根據既有的知識形成自己的分類模型，進行類似結構的分類。決策樹(decision tree)即為一種監督式方法，每個樹枝節點對應到一個位一個以上的屬性，而樹葉節點則代表決策結果。非監督式方法所用來建立模式的資料並不是事先定義好的，而是根據群集方法(clustering)，將資料歸納成不同群組，利用評估技術評估各群之間的關聯性與涵義。許多演算法被應用於資料探勘，如：類神經網路、統計迴歸、群集演算法、基因演算法等，使用何種方法則需視問題類型決定。一般資料探勘包含下列幾項功能：

(1) 分類(classification)：以資料的各種屬性值來判斷該資料的類別。
(2) 估計(estimation)：以資料已知的屬性估計其未知屬性的值。
(3) 預測(prediction)：根據相關資料的變化來預測某一現象是否將要發生。
(4) 相關規律(association rule)：找尋資料屬性間的關係。
(5) 同質分組(clustering)：將資料依其相似程度分成數個群組，其目的是要將組與組之間的差異找出來，同時也要將一個組之中的成員的相似性找出來。

(四)、研究方法

本研究中則應用資料探勘中的同質分組技術，以收集之車流資料分離辨識出不同用路行為並校估參數。研究中假設各種用路行為之車隊於均衡時滿足常態分佈[Helbing, 2001]，如公式(12)：

$$f_e = \frac{1}{\sqrt{2\pi\Theta}} \exp\left(- \frac{\left\| v - u_e \right\|^2}{2\Theta} \right), \tag{12}$$

其中 f_e 為均衡狀態之車流分配，k, u_i and Θ 分別是密度、平均速度與速度變異。而由各種用路行為之分配混合形成總體車流現象，表示如下：

$$F(v) = \sum_{j=1}^{M} \omega_j f\left(x, v, t \middle| \theta_j\right), \tag{13}$$

其中 $F(v)$ 為混合後之分配函數，ω_j 為各駕駛行為之權重，$\sum_{j=1}^{M}\omega_j = 1$。$\theta_j$ 為其參數化之參數。基於以上假設，本研究以同質分組演算法中期望最大(expectation-maximization)法，訓練學習並辨識車流中不同駕駛行為群組。以下將簡要說明同質分組演算法之應用與期望最大法之模式。

同質分組演算法主要群化能夠對資料集內的資料物件進行分堆的動作，當資料集被分成不同的組群後，我們便可得到各組群所突顯的特性，並針對某些感興趣的組群進行更進一步的分析，在資料處理效率的層面上，有很大的幫助；因此，組群化常被用來當作資料分析的前處理步驟或資料挖掘的第一步驟，以利後續的資料探勘工作。

由公式(13)，假設各分配間彼此獨立，則參數 θ 的對數概似函數 (log likelihood) 可表示如下：

$$l\left(\theta \middle| V\right) = \sum_{i=1}^{N} \log \sum_{j=1}^{M} \omega_j f\left(x, v_i, t \middle| \omega_j, \theta_j\right) \text{。} \tag{14}$$

由最大概似法 (maximum likelihood)可知，當 $l\left(\theta \middle| V\right)$ 最大時，可得最佳模式。但因公式(14)中必須對一加總式取對數，$l\left(\theta \middle| V\right)$ 無法直接求得最大值。因此引用另一參數 z 表示 v_i, ω_j 與 θ_j，簡化問題。令 $Z = \{z_i\}_{i=1}^{N}$，且 $z_i = \left(z_{i1}, z_{i2}, \ldots, z_{iM}\right)$，其中 $z_{ij} = 1$ 若且唯若 v_i 由 j 群產生，則資料集合表示成 $V_c = \{V, Z\}$。新的概似函數表示如下：

$$l_c\left(\theta \middle| V, Z\right) = \sum_{i=1}^{N}\sum_{j=1}^{M} z_{ij} \log\left[\omega_j f\left(x, v_i z_i, t \middle| \theta_j\right)\right] \sum_{i=1}^{N}\sum_{j=1}^{M} z_{ij} \log\left[\omega_j f\left(x, v_i, t \middle| z_i, \theta_j\right) f\left(z_i, \theta_j\right)\right] \text{。} \tag{15}$$

式中已不含對數加總式，可進行最大化。但 Z 未知，$l_c\left(\theta \middle| V, Z\right)$ 無法直接計算，以期望最大化法計算，即先計算其期望值，再求其中之最大值，重複迭代計算。令 $Q\left(\theta \middle| \theta_k\right)$ 為其期望值，則期望最大化法可以下列兩步驟表示：

(1) 計算期望值(E-step)：$Q\left(\theta \middle| \theta_k\right) = E\left[l_c\left(\theta \middle| V, Z\right) \middle| X, \theta_k\right]$。

(2) 計算最大值(M-step)：$\theta_{k+1} = \arg\max_{\theta} Q\left(\theta \middle| \theta_k\right)$。

其中 argmax 表示計算使 $Q\left(\theta \middle| \theta_k\right)$ 之參數 θ。上述之期望最大化法為混合模式(mixture model)之期望最大化法之基礎。

根據公式(14)，令 $\theta = \left(u, \Theta\right)$，

$$f_e(x,v,t|u,\Theta)=\frac{1}{\sqrt{2\pi\Theta}}\exp\left(-\frac{\|v-u\|^2}{2\Theta}\right), \tag{16}$$

其概似函數為

$$l_c(\theta|V,Z)=\sum_{i=1}^{N}\sum_{j=1}^{M}z_{ij}\log\left[-\frac{1}{2}\log 2\pi-\frac{1}{2}\log\Theta_j-\frac{1}{2\Theta_j}(v-u_i)^2\right]。 \tag{17}$$

所以其期望值表示為

$$Q(\theta|\theta_k)=\sum_{i=1}^{N}\sum_{j=1}^{M}E[z_{ij}|V,\theta_k]\left[-\frac{1}{2}\log 2\pi-\frac{1}{2}\log\Theta_j-\frac{1}{2\Theta_j}(v-u_i)^2\right] \tag{18}$$

上式中，$E[z_{ij}|V,\theta_k]$為未知，因此問題可簡化為求$E[z_{ij}|V,\theta_k]$，令$h_{ij}^p=E[z_{ij}|V,\theta_k]$，為第 p 次迭代，第 j 個高斯分布中第 i 種速度的機率，可由以下公式求得

$$h_{ij}^p=\frac{f_e(v|u_j^p,\Theta_j^p)}{\sum_{l=1}^{M}f(v|u_l^p,\Theta_l^p)} \tag{19}$$

而最大化步驟則由以下公式求出

$$\frac{\partial E[l_c(\theta|V,Z)|V,\theta_p]}{\partial u_j}=0 \tag{20}$$

$$\frac{\partial E[l_c(\theta|V,Z)|V,\theta_p]}{\partial\Theta_j}=0 \tag{21}$$

其中

$$u_j^{p+1}=\frac{\sum_{i=1}^{N}h_{ij}^p v_i}{\sum_{i=1}^{N}h_{ij}^p} \tag{22}$$

$$\Theta_j^{p+1}=\frac{\sum_{i=1}^{N}h_{ij}^p(v_i-u_j^{p+1})^2}{\sum_{i=1}^{N}h_{ij}^p} \tag{23}$$

根據期望最大化法可求出各高斯分布的權重。然而，在整體混合車流中含有多少種駕駛行為是事先無法確知，因此本研究提出一自動訓練驗證程序計算類別數。步驟如下：

(1) 給定假設駕駛類別總數，如：$M=10$ 種(將高斯混合機率分布記作 GMM - M，如只考慮一種類別則為 GMM – 1；考慮三種類別記作 GMM -3)。

(2) 給定計算迴圈次數 P，設定權重門檻。

(3) 令 $p=1$。

(4) 每次迴圈，隨機將資料按比例分成訓練組與驗證組。

(5) 令 $j=1$。

(6) 以高斯分布配似訓練組車流資料，計算平均值、共變異數與各分布權重。

(7) 以驗證組驗證前一步驟之結果。

(8) 若 $j=M$，檢查 p 是否小於 P 。若等於 P，則停止計算，輸出結果。若 $p<P$，則重回步驟(4)。

(9) 若 $j<M$，則 $j=j+1$，重回步驟(5)。

(10) 判斷類別數，根據權重門檻，接受大於門檻的分布為有效的類別，小於門檻的則捨去。同時也比較 GMM - j 與 GMM - j - 1 兩次高斯混合機率分布的 R-square 值為判斷依據，若前後兩次的值大於 0.95 則將兩次的結果視為相同，以 $j-1$ 為所得的類別數。

(11) 利用所得的分布與權重，校估各種駕駛行為之參數。

因每一種駕駛行為需以一組偏微分方程組進行模擬，用路行為類別越多，所需的方程

組越多，計算越複雜，且須考慮整個系統之之收斂與一致性(self-consistent)。為顧及模擬時之可行性，採用權重門檻忽略影響較小的用路行為，簡化所需計算的方程組。

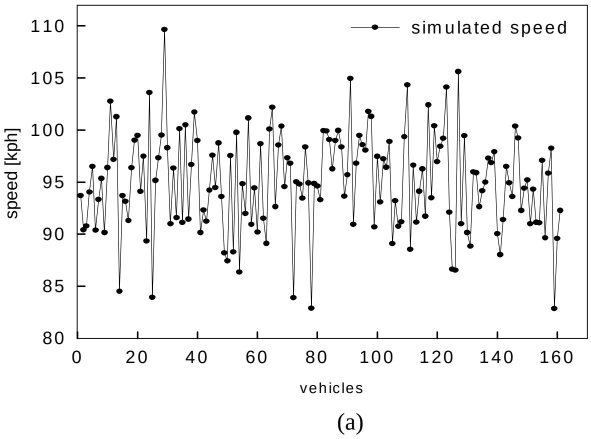
(五)、結果討論

本研究將以兩簡例驗證方法之可行性。分別以一個高斯分布與兩個高斯分布產生速度之隨機變數，其模擬的參數與以資料探勘判別出來之結果，如表 1 所示。簡例一中，以一個平均速度為 95 (kph)，變異數為 25 的高斯分布隨機產生 160 組速度資料，如圖 2(a)所示。以辨識演算法學習判別，可以得到資料可以用一個高斯分布(GMM –1)描述，也就是屬於一種用路行為，如圖 2(b)所示。因為用兩個高斯分布(GMM – 2)學習後所得的結果與 GMM –1 沒有差異，兩者間的 R – square 值為 0.98。因此，可得到簡例一中只有一個用路行為的結論，與模擬時的假設一致。

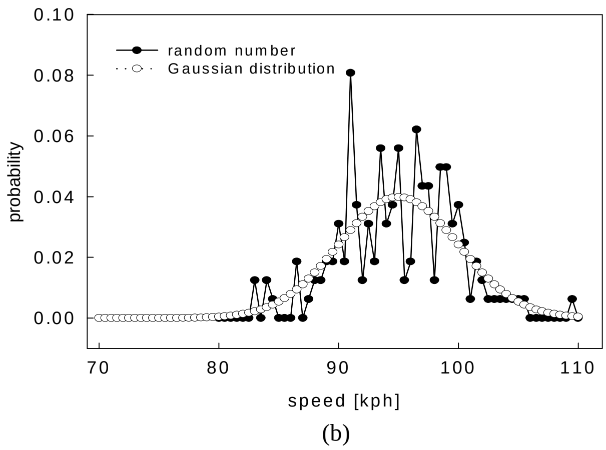
簡例二中，以高斯分布各隨機產生 160 組速度資料，如表 1 與圖 3(a)所示。以辨識演算法學習判別，可以得到資料可以用二個高斯分布(GMM – 2)描述最佳，也就是屬於二種用路行為，如圖 3(b)與(c)所示，因為以 GMM – 1 所得的結果無法看出兩個尖峰型態，而 GMM – 3 學習所得的結果與 GMM – 2 沒有差異，兩者間的 R – square 值為 0.99。因此，可得到簡例二中包括二種用路行為，與模擬時的假設一致。

表 1. 數值例參數與辨識結果

	平均數	變異數	權重
Case 1	95	25	-
GMM –1	94.98	22.18	-
GMM –2, distribution 1	92.25	15.65	0.504
GMM –2, distribution 2	97.76	13.49	0.496
Case 2, distribution 1	80	4	0.5
Case 2, distribution 2	90	36	0.5
GMM –1	85.06	43.21	-
GMM –2, distribution 1	80.41	5.02	0.558
GMM –2, distribution 2	90.94	29.58	0.442
GMM –3, distribution 1	80.13	4.17	0.525
GMM –3, distribution 2	93.63	21.07	0.271
GMM –3, distribution 3	86.37	11.12	0.204

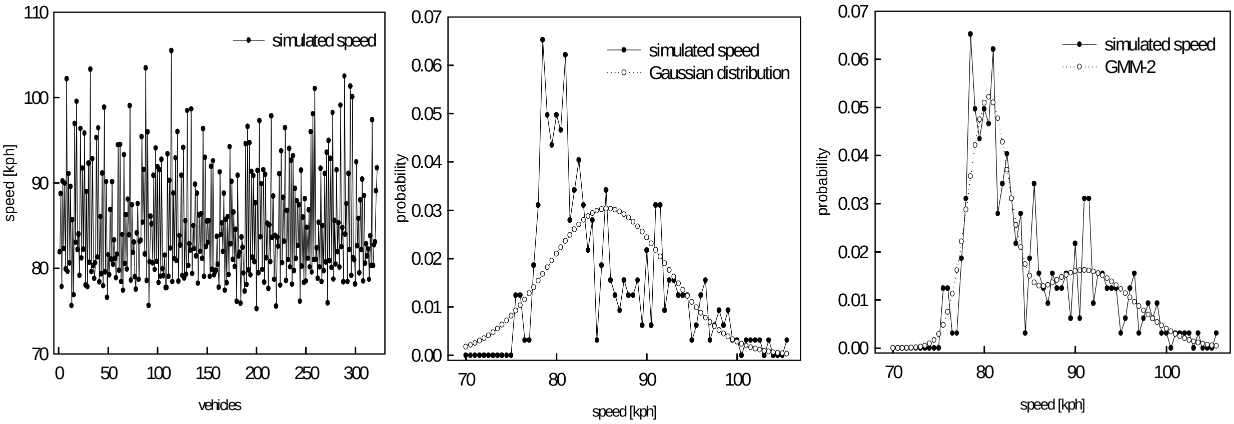


(a)



(b)

圖 2(a)簡例一中，隨機產生之速度分布；(b)簡例一之學習辨識結果



(a) (b) (c)

圖 3 簡例二中，(a)隨機產生之速度分布；(b) GMM –1 學習辨識結果；(c)GMM – 2 學習辨識結果

參考文獻

Adriaans, P. and D. Zantinge (1996), Data mining, Addison-Wesley Inc.

Cho, H. J., and S. C. Lo (2002), Modeling of Self-consistent Multi-class Dynamic Traffic Flow Model, Physica A, pp. 342~362.

Fayyad, U. M. et al., Ed. (1996), Advances in knowledge discovery and data mining, MIT Press.

Freitas, A. A. (2002), Data mining and knowledge discovery with evolutionary algorithms, Springer-Verlag.

Han, J. and M. Kamber (2001), Data mining: concepts and techniques, Morgan Kaufmann Publishers.

Hand, D., H. Mannila and P. Smyth (2001), Principles of data mining, MIS Press.

Helbing, D. (1995), Improved Fluid-Dynamic Model for Vehicular Traffic, Physical Review E, Vol. 51, No. 4, pp.3164-3169.

Helbing, D. (1996), Derivation and Empirical Validation of a Refined Traffic Flow Model, Physica A, Vol. 233, pp.253-282.

Helbing, D. (1997), Empirical Traffic Data and Their Implications for Traffic Modeling, Physical Review E, Vol. 55, pp. R25-R28.

Helbing, D. (2001), MASTER: Macroscopic traffic simulation based on a gas-kinetic, non-local traffic model, Transportation Research Part B, Vol. 35, pp. 183-211.

Hoogendoorn, S. P. and P. H. L. Bovy (1998), Modeling Multiple User-Class Traffic, Transportation Research Record, Vol. 1644, pp.57-70.

Hoogendoorn, S. P. (1999), Multiclass Continuum Modelling of Multilane Traffic Flow, Doctoral Dissertation, University of Delft.

Hoogendoorn, S. P. and P. H. L. Bovy (2000), Continuum modeling of multiclass traffic flow, Transportation Research Part B, Vol. 34, pp. 123-146.

Lighthill, M. J., and G. B Whitham (1955), On Kinematics Waves II. A Theory of Traffic Flow on Long Crowded Road, London, Proceedings Royal Society, A229, pp.317-345.

Lo, S.-C. (2002), Modeling and simulation of vehicular kinetic flow – from the viewpoint of Boltzmann transport equation, Ph.D. Thesis, National Chiao Tung University.

Michalopoulos, P. G., G. Stephanopoulos, and V. B. Pisharody (1980), Modeling of Traffic Flow at Signalized Links, Transportation Science, Vol. 14, No. 1, pp.9-41.

Michalopoulos, P. G., and V. Pisharody (1980), Plaoon Dynamics on Signal Controlled Arterial, Transportation Science, Vol. 14, No. 4, pp.365-396.

Michalopoulos, P. G., G. Stephanopoulos, and G. Stephanopoulos (1981), An Application of Shock Wave Theory to Traffic Signal Control, Transportation Research Part B, Vol. 15, No. 1, pp.35-51.

Michalopoulos, P., D. Beskos and Y. Yamauchi (1984), Multilane Traffic Flow Dynamics: Some Macroscopic Consideration, Transportation Research Part B, Vol. 18, pp. 377-393.

Michalopoulos, P. G., P. Yi, and A. D. Lyrintzis (1993), Continuum Modelling of Traffic Dynamics for Congested Freeways, Transportation Research Part B, Vol. 27, No. 4, pp.315-352.

Payne, H. J. (1979), Freflo: A Macroscopic Simulation Model of Freeway Traffic, Transportation Research Record, Vol.722.
Richards, P. I. (1956), Shock Waves on the Highway, Operation Research, Vol. 4, No. 1, pp.42-51.
Roiger, R. J. and M. W. Geatz (2003), Data mining — A tutorial-based primer, Pearson Education, Inc.
Trueblood, R. P. and J. N. Lovett Jr. (2001), Data mining and statistical analysis using SQL, Apress Inc.
Westphal, C. and T. Blaxton (1998), Data mining solutions: methods and tools for solving real-world problems, Wiley Inc.

三、計畫成果自評

本研究以資料探勘中之同質分組技術為基礎，提出一由交通資料中萃取出不同駕駛行為類別之方法。經數值例驗證，此一技術可辨識出多用路行為車流中之不同用路行為，較傳統之調查方式節省時間與金錢成本，**完全符合**計劃書所提的內容。本研究部分成果**已發表**於國際研討會(2007 International Conference of Computational Methods in Sciences and **Engineering, ICCMSE 2007)，共發表兩篇論文，發表文章資訊如下:**

Shih-Ching Lo, "Classification of Driving Behavior by Pattern Recognition in Multiclass Users Traffic Flow," presented in International Conference of Computational Methods in Sciences and Engineering, Corfu, Greece, Sept. 25-30, 2007.

Shih-Ching Lo, Hsiao-Wei Tsai, Hsiao-Hui Wang and Meng-Hsi Chen, "The Effect of Driving Behavior on Multiclass Users Traffic Flow," presented in International Conference of Computational Methods in Sciences and Engineering, Corfu, Greece, Sept. 25-30, 2007.

文章如附件一與二所示。此外，結果將延伸應用並投至相關領域之國際學術期刊。

附件一

Classification of Driving Behavior by Pattern Recognition in Multiclass Users Traffic Flow

Shih-Ching Lo

Department of Transportation Technology and Logistics Management, Chung Hua University, No. 707, Sec. 2, WuFu Rd., Hsinchu, 300, Taiwan

Abstract. Understanding driving behavior is a complicated researching topic. To describe accurate speed, flow and density of a multiclass users traffic flow, an adequate model is needed. Mostly, user's classes are determined by types of vehicles. However, it is unrealistic to consider drivers with the same type of vehicles have the same driving behavior. Conventionally, classifying driving behavior is obtained through tracking trace of individual vehicles, experimenting by driving simulator or inquiring by questionnaire. It costs a lot and may produce bias because of the design of questionnaire or experiment. Therefore, a new method, which is based on pattern recognition technique, is proposed to classify driving behavior in multiclass user traffic flow. In this study, driving behavior, which performs as speed distributions, is assumed to be Gaussian distributions. According to the assumption, the expectation-maximization algorithm is employed to train and classify different driving behavior. With the method, a economical and automatic way for traffic data processing and parameter extracting is obtained.

Keywords: traffic flow, pattern recognition, classification, multiclass users traffic flow..
PACS: 05.20.Dd, 51.10.+y, 89.40.-a, 89.40.Bb.

INTRODUCTION

With the rising demand of automobile and highway usage in recent years, traffic congestion in metropolitans causes great economical loss and pollution. Traffic flow theory provides the description of the fundamental traffic flow characteristics and analytical techniques to draw up control strategies so as to improve the performance of road systems. In the real world, traffic flow is heterogeneous; that is, there are different types of vehicles and different driving behavior on a road. In order to improve traffic conditions on roads, gaining a clear insight into the behavior of the heterogeneous traffic flow is important [1-6]. For convenience's sake, driving behavior is defined as users' classes, which are determined by types of vehicles, such as buses, trucks, cars or motorcycles. However, drivers with the same types of vehicles may have different driving behavior in reality. On the other hand, driving behavior is studied through inquiring drivers by interview, telephone, mail or web page, investigating by driving simulator or tracking trace of individual vehicles. It costs a lot and is time-consuming. Also, it may produce perceptual bias because of the design of questionnaire or experiment. Therefore, a pattern recognition based technique is proposed to classify driving behavior in multiclass traffic flow in this study.

Pattern recognition is based on the observation of past experience or knowledge. Today, useful applications of automatic pattern recognition are prevalent. As computers and the methods of automatic pattern recognition progress, more and more fascinating applications are being discovered in fields as broad as finance, manufacturing, and medicine. Generally, speed distribution of a road is considered as the performance of driving behavior and is examined as a Gaussian distribution empirically [7-8]. Based on the assumption, speed distribution of a multiclass users traffic flow can be considered as mixture of multiple Gaussian distributions [9-11]. If we can recognize how many Gaussian distributions are included in the mixture speed distribution, we can identify the number of user-class on the road. Therefore, an expectation-maximization based pattern recognition method for multiclass traffic flow is proposed in this study. According to the method, users' classes are identified by speed data successfully.

PATTERN RECOGNITION

In this study, expectation-maximum algorithm (EM algorithm) based pattern recognition method is proposed. With the method, parameters of multiclass traffic flow model can be obtained by collected speed data directly. Firstly, we assume the speed data is denoted as $V = \{v_i\}_{i=1}^N$. According to Helbing [7-8], the equilibrium speed of each user-class can be considered as a Gaussian distribution, which is

$$f_j(v_i) = \frac{1}{\sqrt{2\pi\Theta_j}} \exp\left(-\frac{\|v_i - u_{ej}\|^2}{2\Theta_j} \right), \tag{1}$$

where f_j is equilibrium distribution of user-class j, v_i is individual speed, u_{ej} is mean speed of user-class j, and Θ_j is speed variance of user-class j. Thus, the whole speed distribution of traffic flow is given by

$$F(v_i) = \sum_{j=1}^{M} \omega_j f\left(x, v_i, t \middle| \theta_j\right), \tag{2}$$

where f_j in Eq. (1) is parameterized by θ_j, ω_j is the weight of user-class j of the mixture and $\sum \omega_j = 1$, $j = 1, \ldots, M$. The log likelihood of the parameters is written as

$$l(\theta|V) = \sum_{i=1}^{N} \log \sum_{j=1}^{M} \omega_j f\left(x, v_i, t \middle| \omega_j, \theta_j\right). \tag{3}$$

By the maximum likelihood principle, the best model of the data has the parameters that maximize $l(\theta|V)$. Unfortunately, $l(\theta|V)$ cannot be easily maximized because it involves a logarithms of a sum. Therefore, another parameter z is introduced to replace θ_j and ω_j so as to simplify the problem. z indicates that the speed data belongs to which user-class. Let $Z = \{z_i\}_{i=1}^N$, where $z_i = (z_{i1}, z_{i2}, \ldots, z_{iM})$ and $z_{ij} = 1$ iff v_i belongs to user-class j. The new data set is denoted as $V_c = \{V, Z\}$ and the new log likelihood function is rewritten as

$$l_c(\theta|V,Z) = \sum_{i=1}^{N}\sum_{j=1}^{M} z_{ij} \log\left[\omega_j f\left(x, v_i z_i, t \middle| \theta_j\right)\right] \sum_{i=1}^{N}\sum_{j=1}^{M} z_{ij} \log\left[\omega_j f\left(x, v_i, t \middle| z_i, \theta_j\right) f\left(z_i, \theta_j\right)\right], \tag{4}$$

which does not involve a logarithms of a sum. However, Z is unknown, $l_c(\theta|V,Z)$ cannot be utilized directly. We replace $l_c(\theta|V,Z)$ with its expectation $Y(\theta|\theta_k)$. According to previous studies [9-11], $l_c(\theta|V,Z)$ can be maximized by the following two steps:

(1) E-step: $$Y(\theta|\theta_k) = E\left[l_c(\theta|V,Z) \middle| X, \theta_k\right], \tag{5}$$

(2) M-step: $$\theta_{k+1} = \arg\max Y(\theta|\theta_k), \tag{6}$$

where argmax denotes finding the parameter θ that maximize $Y(\theta|\theta_k)$. The E-step calculates the expectation of the speed data log likelihood, and the M-step finds the parameters that maximize this likelihood. These two steps form the basis of the EM algorithm for mixture model. From Eqs (1), (3)~(6), let $\theta = (u, \Theta)$, the explicit form of likelihood function is written as

$$l_c(\theta|V,Z) = \sum_{i=1}^{N}\sum_{j=1}^{M} z_{ij} \log\left[-\frac{1}{2}\log 2\pi - \frac{1}{2}\log\Theta_j - \frac{1}{2\Theta_j}(v_i - u_{ej})^2\right]. \tag{7}$$

The expectation of E-step is

$$Y(\theta|\theta_k) = \sum_{i=1}^{N}\sum_{j=1}^{M} E\left[z_{ij}|V,\theta_k\right]\left[-\frac{1}{2}\log 2\pi - \frac{1}{2}\log\Theta_j - \frac{1}{2\Theta_j}(v_i - u_{ej})^2\right]. \tag{8}$$

In Eq. (8), $E\left[z_{ij}|V,\theta_k\right]$ is unknown. Therefore, the problem is simplified to solve the unknown term $E\left[z_{ij}|V,\theta_k\right]$. Let $h_{ij}^p = E\left[z_{ij}|V,\theta_k\right]$ be the probability of ith speed, which belongs to the jth Gaussian distribution in the pth iteration. h_{ij}^p is computed by

$$h_{ij}^p = f_j\left(v \middle| u_j^p, \Theta_j^p\right) \Big/ \sum_{l=1}^{M} f\left(v \middle| u_l^p, \Theta_l^p\right). \tag{9}$$

Next, the M-step is computed by

$$\partial E[l_c(\theta|V,Z)|V,\theta_p]\Big/\partial u_j = 0\,, \tag{10}$$

$$\partial E[l_c(\theta|V,Z)|V,\theta_p]\Big/\partial \Theta_j = 0\,, \tag{11}$$

where $u_j^{p+1} = \sum_{i=1}^{N} h_{ij}^p v_i \Big/ \sum_{i=1}^{N} h_{ij}^p$ and $\Theta_j^{p+1} = \sum_{i=1}^{N} h_{ij}^p \left(v_i - u_j^{p+1}\right)^2 \Big/ \sum_{i=1}^{N} h_{ij}^p$.

According to the EM algorithm, we can obtain the weight of each Gaussian distribution and the number of user-class.

NUMERICAL RESULTS AND DISCUSSION

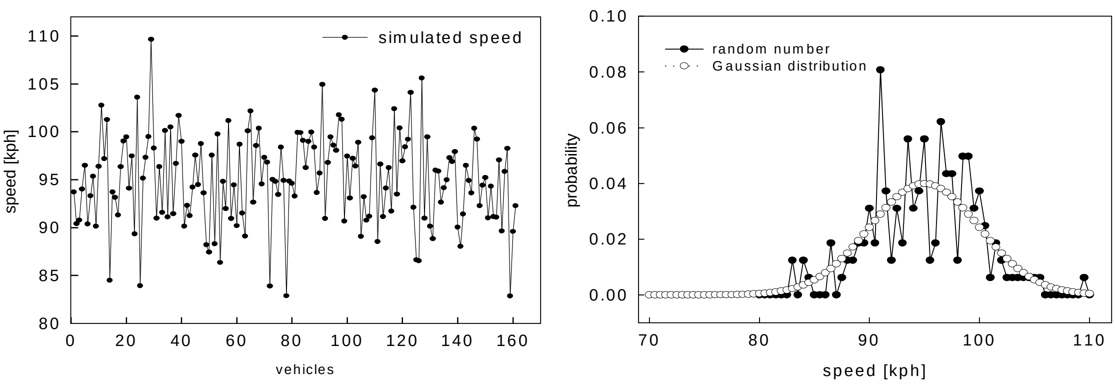
Two numerical examples are employed to verify the method. The speed data of case 1 is generated by single Gaussian distribution and case 2 is generated by mixing two Gaussian distributions stochastically. Case 1 includes 160 data points and case 2 includes 320 data points. The simulated scenario and results are given in Table 1.

In case 1, P and M are given as 5 and 2, respectively. Figure 1 (a) illustrates the generated speed and (b) is the comparison of GMM-1 and the generated data. By the procedure presented in previous section, GMM-1 fits the generated data well. From Table 1, GMM-1 has a good agreement with the generated data. Both mean speeds are almost the same while variances have a little difference. At the same time, the R-square of GMM-1 and GMM-2 is 0.98, that is, there is no significant difference between GMM-1 and GMM-2. Thus, we can conclude that only one user-class in case 1, which is the same as the generated distribution.

In case 2, P and M are given as 5 and 3, respectively. Figure 2 (a) illustrates the generated speed and (b) is the comparison of GMM-1 and the generated data. Figure 2 (c) shows the comparison of GMM-2 and the generated data. In this case, GMM-3 fits the generated data best. However, the R-square of GMM-2 and GMM-3 is 0.99; that is, there is no significant difference between GMM-2 and GMM-3. Also, GMM-2 has a good agreement with the generated data according to Table 1. The mean speed of GMM-2 is almost the same as the generated data while variance and weight of GMM-2 have a little difference. Hence, we can conclude that there are two user-classes in case 2, which is the same as the generated distribution.

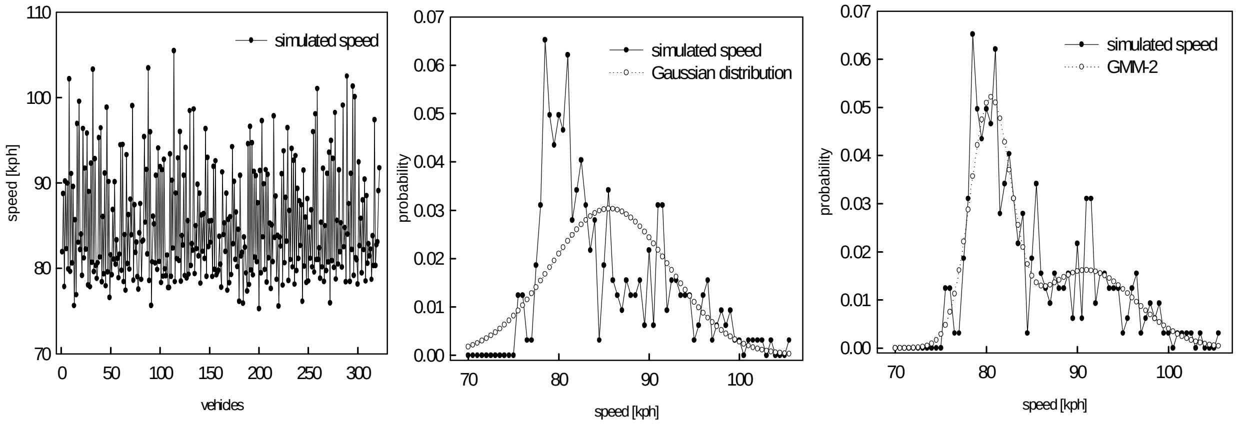
Table 1. Simulated scenario and classified results

	mean	**variance**	**weight**
Case 1	95	25	-
GMM -1	94.98	22.18	-
GMM -2, distribution 1	92.25	15.65	0.504
GMM -2, distribution 2	97.76	13.49	0.496
Case 2, distribution 1	80	4	0.5
Case 2, distribution 2	90	36	0.5
GMM -1	85.06	43.21	-
GMM -2, distribution 1	80.41	5.02	0.558
GMM -2, distribution 2	90.94	29.58	0.442
GMM -3, distribution 1	80.13	4.17	0.525
GMM -3, distribution 2	93.63	21.07	0.271
GMM -3, distribution 3	86.37	11.12	0.204



(a) (b)

FIGURE 1. (a) speed data generated by single Gaussian distribution; (b) comparison of generated data (denoted by random number) and GMM-1 (denoted by Gaussian distribution).



(a) (b) (c)

FIGURE 2. (a) speed data generated by two Gaussian distributions; (b) comparison of generated data (denoted by random number) and GMM-1 (denoted by Gaussian distribution); (c) comparison of generated data and GMM-2.

CONCLUSIONS AND PERSPECTIVES

In this study, an EM algorithm based pattern recognition method for multiclass traffic flow is presented and verified by two numerical examples. This method can extract parameters of multiclass users by speed data directly, which saves time and money. Since speed data, which can be collected by traffic surveillance systems, is the only necessary input, it is possible to classify user-class and extract parameters automatically. Furthermore, the method takes computational complexity of traffic flow simulation into account by setting threshold of weight and comparison of GMM models. The two considerations minimize the number of user-class without losing feasibility. Therefore, an integrated multiclass traffic control system is achieved by coupling our method with a multiclass traffic flow model.

ACKNOWLEDGMENTS

The work was partially supported by the National Science Council (NSC), Taiwan, under Contract NSC 96-2415-H-216-001.

REFERENCES

1. D. Helbing, *Trans. Res.*, **35B**, 183-211 (2001).
2. S. P. Hoogendoorn, and P. H. L. Bovy, *Trans. Res.*, **34B**, 123-146 (2000).
3. P. Bagnerini, and M. Rascle, *SIAM J. Math. Anal.*, **35**, 949-973 (2003).
4. C. M. J. Tampère, "Human-Kinetic Multiclass Traffic Flow Theory and Modelling (With Application to Advanced Driver Assistance Systems in Congestion)", Ph.D. Thesis, Delft University of Technology (2004).
5. D. Ngoduy, "Macroscopic Discontinuity Modeling for Multiclass Multilane Traffic Flow Operations", Ph.D. Thesis, Delft University of Technology (2006).
6. T.W. Schaap, and B. van. Arem,: A Comprehensive Driver Behavior Model for the Evaluation of Intelligent Intersections, Proceeding of 13th World Congress on ITS, London, UK (2006).
7. D. Helbing, *Physica A*, **233**, 253-282 (1996).
8. D. Helbing, *Phys. Rev. E*, **55**, R25-R28 (1997).
9. A. P. Dempster, N. M. Laird and D. B. Rubin, *J. R. Stat. Soc. B*, **39**, 1–22 (1977).
10. E. Redner and H. Walker, *SIAM Rev.*, **26**, 195-239 (1984).
11. J. Bilmes, A Gentle Tutorial on the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models, Technical Report, University of Berkeley, ICSI-TR-97-021, (1997).

附件二

The Effect of Driving Behavior on Multiclass Users Traffic Flow

Shih-Ching Lo, Hsiao-Wei Tsai, Hsiao-Huei Wang and Mong-Xi Chen

Department of Transportation Technology and Logistics Management, Chung Hua University, No. 707, Sec. 2, WuFu Rd., Hsinchu, 300, Taiwan

Abstract. Complex traffic system seems to be simulated successfully by cellular automaton (CA) models. Various models are developed to understand single-lane traffic, multilane traffic, lane-changing behavior and network traffic situations based on the basic CA rules proposed by Nagel et al. In this paper, a multi-class user traffic flow CA model is proposed to investigate the influence of driving behavior in traffic flow. Slow down possibility and maximal speed are two main variables, which determine driving behavior. Simulation scenario shows that the diversity of driving behavior will induce unstable traffic flow even chaos phenomena. Traffic controlling and management strategies are also discussed in this study. According to the results, optimal strategies may be developed and maximize traffic flow.

Keywords: traffic flow, multiclass users traffic flow, driving behavior, cellular automation.
PACS: 89.40.-a, 89.40.Bb, 02.60.Cb

INTRODUCTION

Today, there are many indications of the complexity of living in the world. One of them is the road using behavior. As the trend of increasing travel demand, planning, design, prediction, control and management of the transportation system become more and more important. Traffic flow theory provides the description of the fundamental traffic flow characteristics and analytical techniques. In the research of traffic flow, simplified models have been proposed and these models still capture the essentials of the dynamics of the transportation system. Cellular automation (CA) is one of these models. Although the concept of CA is first proposed long ago[1], CA has begun to receive wide attention of statistical physics community only after the simple formulation by Nagel and Schreckemberg [2]. In CA, a road is represented as a string of cells, which are either empty or occupied by exactly one vehicle. Movement takes place by hopping between cells.

Due to the simplicity of computation, CA has been generalized to signalized intersection, multilane multiclass traffic flow [3-7], inhomogeneous mixed traffic flow [8] and large traffic networks. Nagel [3] compared the other models with CA and made some conclusions as follow:

(1) Robust computing: CA is known to be numerically robust especially in complex geometries.
(2) University: Intuitively, a relatively simple microscopic model should be able to show the essential features of traffic jams. One might even speculate that the critical exponents of traffic jam formation are universal.
(3) Towards minimal models: The present results show that close-up vehicle-following behavior is not the most important aspect to traffic model. The important crucial aspect is to model deviations from the optimal (smooth) behavior and the ways in which they lead to jam formation. Another important aspect is the acceleration behavior, that mostly determines the maximum flow out of a jam.
(4) Traffic dynamics: Fast running and easy to implement CA can be very useful in interpreting measurements.
(5) Microscopic simulation: CA is inherently microscopic, which allows one to add individual properties to each vehicle.
(6) Stochastic and fluctuations: Last but not least, CA is stochastic in nature; thus, different results may be produced by using different random seeds even when the simulation is starting from identical initial

conditions. The traffic system is inherently stochastic and the variance of the outcomes is an important variable itself.

The simulation results were compared with data extracted from real traffic system in the USA and Germany [5-6]. Verification of CA-models on German and American motorways and urban traffic networks shows fairly realistic results on a macroscopic scale. In this study, we proposed a modified CA procedure and applied the procedure to multiclass users traffic flow. Furthermore, analysis of multiclass users traffic flow is presented based on the simulation results.

CELLULAR AUTOMATION OF MULTICLASS USERS TRAFFIC FLOW

CA-models describe the traffic system as a lattice of cells of equal size (typically 7.5m). A CA-model describes the movements of vehicles from cell to cell in a discrete way [3-4]. The size of the cell is chosen to be equal to the velocity of vehicle that moves forward one cell during one time step. The vehicle's velocity can only assume a limited number of discrete values ranging from zero to *v*max. The process can be split-up into four steps:

(1) *Acceleration*. If time step is less than total simulation time, let each vehicle with velocity be smaller than its maximum velocity *v*max, accelerate to a higher velocity, i.e. $v = \min(v\max, v+1)$.

(2) *Deceleration*. If the velocity is smaller than the distance gap d to the preceding vehicle (v'), the vehicle will decelerate: $v = \min(v, d)$.

(3) *Dawdling*. With given slow-down probability p, the velocity of a vehicle decreases spontaneously: $v = \max(v-1, 0)$.

(4) *Propagation*. Let each vehicle move forward v cells and let time step increase one. Then, repeat the procedure: acceleration, deceleration, dawdling and propagation.

In this study, we assume that if the velocity is larger than the distance gap to the preceding vehicle and the velocity is larger than the velocity of the preceding vehicle, the following vehicle will decelerate to keep the velocity of the preceding vehicle (i.e., $v = v'$). Otherwise, the following vehicle will keep its velocity. Therefore, step (2) should be modified as (2') and an additional step (3-1) should be inserted between steps (3) and (4). The modified process is given as

(2') *Deceleration*. If $v > d$, then check if $v > v'$ or not. If the answer is no, keep the velocity the same. If the answer is yes, let $v = v'$.

(3-1) *Deceleration:* Repeat step (2').

According to the process, driving behavior is determined by two parameters, *v*max and p; that is, maximum velocity and slow-down probability. Different behavior can be simulated by different *v*max and p.

RESULTS AND DISCUSSION

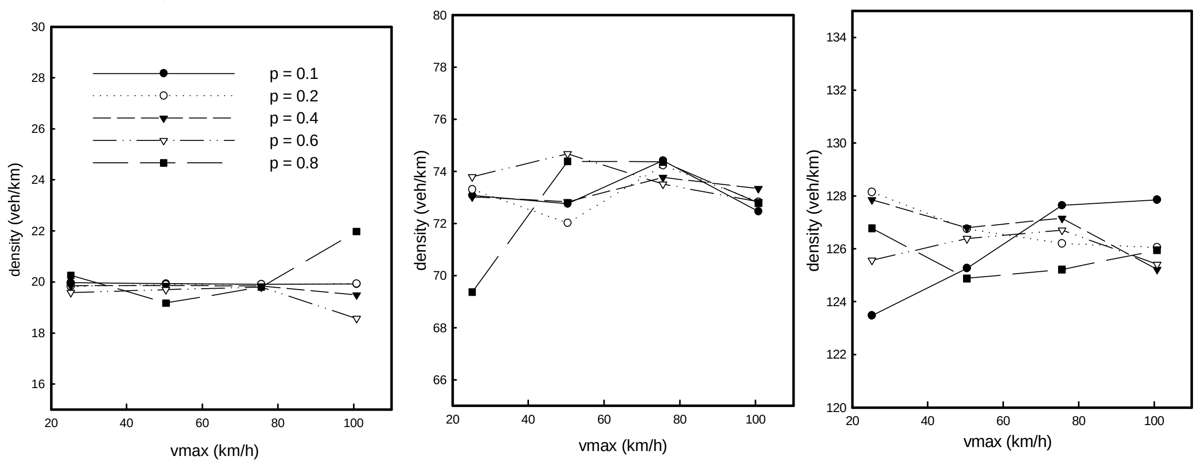
The rules proposed previously will be used throughout the paper, with different simulated scenario. Typically, the length of a cell was taken as 7.5 m, time step is 1 second, *v*max is 5 (i.e., 135 km/h). In Taiwan, the upper speed limit of No. 1 National Freeway is 100 km/h. Therefore, the length of a cell is considered as 7 m, the maximum *v*max is 4 (i.e., 100.8 km/h). All simulations are performed in a single lane circle of length 1.5km (i.e., 214 cells). Density is estimated every 30 seconds. 3,600 steps are simulated. Simulated number of vehicles on the road varies from 10 to 190, *v*max varies from 1 to 4 (i.e., 25.2 km/h to 100.8 km/h), slow-down probability (p) varies from 0.1 to 0.9. Single user traffic flow is simulated first. Parts of the results are illustrated in Figs. 1 and 2. The simulated numbers of vehicles, which is denoted by N, are 30, 110 and 190, which imply the average normalized densities on the whole road are 0.15 (free flow), 0.51 (intermediate flow) and 0.883 (congested flow). Figures 1 and 2 show the variation of density and volume with *v*max, respectively. The data point is the mean of one hour. Therefore, density looks smooth. The variation can be observed by variance of density. Since volume is equal to density multiply to speed, the fluctuation of speed is similar to the fluctuation of volume while density is smooth. According to the figures and the results, when traffic is in the regime of free flow (N is small), the variation of density and volume increase with *v*max and p increase. Larger *v*max allows higher speed and larger p implies more vehicles may decelerate in free flow; i.e., large *v*max and p induce unstable traffic in free flow regime. Since larger *v*max allows higher speed, it also induces larger volume in free flow regime. On the other hand, the same phenomena cannot be observed in intermediate and congested flow. Because drivers cannot drive freely when the number of vehicles on the road increases. Interaction among vehicles decrease the mean speed and volume, whereas increases density on the road. This result can be observed obviously in Fig. 1 (b) and 2 (b). Figures 1 (c) and 2 (c) present an interesting result; i.e., if driving behavior is stable (p is small), the volume on the road is still high. This observation gives a good reason to develop an automatic highway system. Generally, the mean density is not sensitive to *v*max and *p* as speed and volume are. While traffic is in free flow regime, speed and volume are dominated by *v*max and *p*. On the other hand, speed and volume are dominated by *p* while traffic is in intermediate and congested regime. Most of the CA traffic simulation considered *p* to be a constant (mostly *p* = 0.25). However, in this study we found *p* should depend on density. That is, when density is high, the slow-down probability is large and vice versa. According to the results, *p* is suggested to be equal to reciprocal of normalized density (or so-called dimensionless density, denoted by *k'*). The comparing result of fixed *p* = 0.2 and *p* = *k'* are illustrated in Fig. 3. The speed-density relation simulated by CA is similar to real data [5-6]. By Fig. 3, the trends of both curves are the same and the variance, which determined by *p* are different. *p* = 0.2 shows a uniform variance of speed-density relation. *p* = *k'* shows the intermediate flow has the largest variance and the variance of the congested flow is relative small. Therefore, *p* = *k'* is more realistic than given *p* as a constant.

Driver behavior is assumed to be determined by *v*max and *p* in this study. Each class of users moves forward by the rules mentioned above. *v*max in multiclass users traffic flow cellular automation simulation is considered as the minimum of desired speed of each class and speed limit of the road. *p* is assumed to be close to 1/*k*' and different class of user may have different *p*. In this study, a two-class users traffic flow is simulated. Firstly, we keep *v*max the same and vary *p*. If most of the drivers have large *p* and the others have small *p*, density increases, whereas speed and volume decreases. Then, we keep *p* the same and vary *v*max. If difference of *v*max between two kinds of drivers is large, density increases, whereas speed and volume decreases. In both cases, the multiclass users traffic flow becomes unstable and is dominated by slow vehicles. The influence of *v*max is larger than *p* in free-flow regime. According to the results, the proportion of each user class is also an important factor. If the proportion of slow vehicles is large, the speed and volume will be much less than single user traffic flow.

Mostly, speed is dominated by the slowest vehicles, especially in free and intermediate flow. Different driving behavior may increase variance and decrease level of service. Since *v*max, *p* and proportion of each user class are three factors of multiclass users traffic flow, several strategies can be applied to stable traffic flow. A long-term strategy is to enhance driving training and education so as to unify driving behavior. Setting lower speed limit is also a control strategy to stabilize traffic flow. Of course, developing the automatic highway system or the automatic vehicle control system can ensure the driving behavior being unified.

CONCLUSION

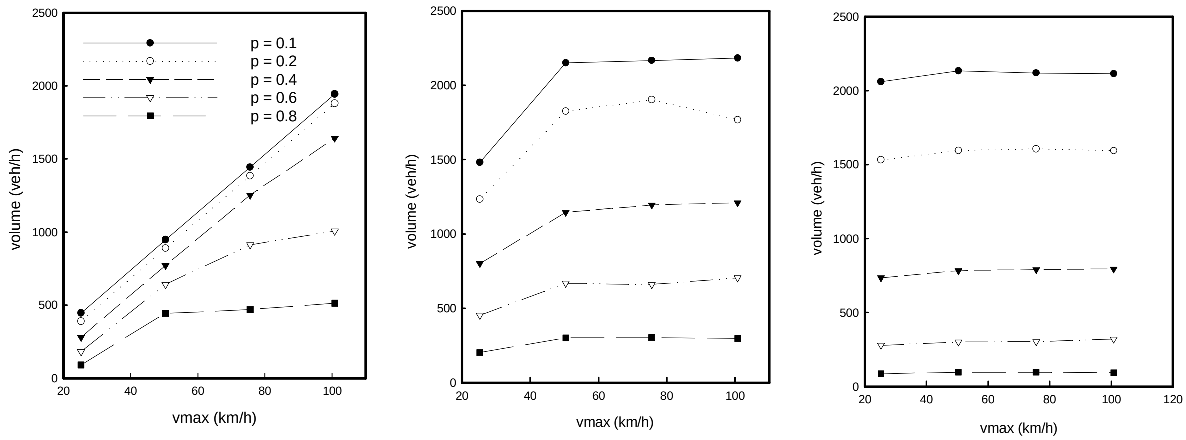
In multiclass users traffic flow, *v*max, *p* and the proportion of each class users are three dominated factors. Variances of speed and volume in multiclass users traffic flow are larger than variances of speed and volume in single user traffic flow. These results are interesting. However, before final conclusions can be stated, a lot more work has to be done to confirm these results, such as more simulation and comparison of simulation results and empirical data. In addition, there are still other factors, which influence multicalss users traffic flow, such as speed-up probability, acceleration and deceleration ability. These will be left for further studies.



(a) (b)

(c)

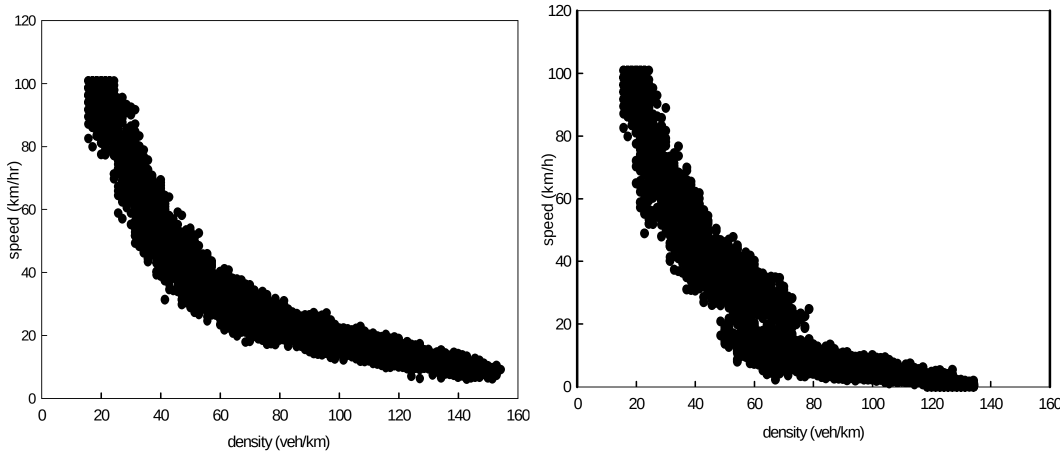
FIGURE 1. Mean density of different *v*max and slow-down probability for (a) *N* = 30; (b) *N* = 110 and (c) *N* = 190.



(a) (b)

(c)

FIGURE 2. Mean volume of different *v*max and slow-down probability for (a) N = 30; (b) N = 110 and (c) N = 190.



(a) (b)

FIGURE 3. Comparison of speed-density distribution of (a) fixed p = 0.2 and (b) $p = k'$ with *v*max = 4.

ACKNOWLEDGMENTS

The work was partially supported by the National Science Council (NSC), Taiwan, under Contract NSC 96-2415-H-216-001.

REFERENCES

1. D. L. Gerlough, "Simulation of Freeway Traffic by An Electronic Computer", Proceedings of 35th Annual Meeting of Highway Research Board, Washington, D. C. (1956).
2. K Nagel and M, Schreckenberg *J. Physique I* **2** 2221-2229 (1992).
3. K. Nagel, *Phys. Rev. E* **53**, 4655-4672 (1996).
4. K. Nagel, D. E. Wolf, P. Wagner and P. M. Simon. , *Phys. Rev. E*, **58** No. 2, (Aug. 1998)
5. J. Esser, L. Neubert, J. Wahle, and M. Schreckenberg, Microscopic online Simulations of Urban Traffic. In Ceder, A. (ed), Proceedings of the 14th International Symposium of Transportation and Traffic Theory, Jerusalem, 517-534 (1999).
6. N. Wu and W. Brilon, Cellular Automata for Highway Traffic Flow Simulation. In: Ceder, A. (ed), Proceedings 14th International Symposium on Transportation and Traffic Theory (Abbreviated presentations), 1-18 (1999).
7. Daganzo C.F., *Trans. Res. B* **28**, 269 – 287 (1994).
8. L. W. Lan and C. W. Chang, Inhomogeneous cellular automata modeling for mixed traffic with vehicles and motorcycles, *J. Adv. Trans.*, **39**, 323-349 (2004)

出席國際學術會議心得報告

計畫編號	NSC 96－2415－H－216－001
計畫名稱	以資料探勘技術辨識混合車流中駕駛行為類別之研究
出國人員姓名 服務機關及職稱	羅仕京，中華大學運輸科技與物流管理學系，助理教授
會議時間地點	2007/09/25~2007/09/30，Corfu，Greece(希臘)
會議名稱	2007 International Conference of Computational Methods in Sciences and Engineering (2007 ICCMSE)
發表論文題目	Classification of Driving Behavior by Pattern Recognition in Multiclass Users Traffic Flow

一、參加會議經過

ICCMSE 由創辦開始至今年為第五屆，研討會旨在促進數學計算方法在科學與工程上之應用。此次參加共發表二篇論文，題目分別為發表於 9/25 的 “Classification of Driving Behavior by Pattern Recognition in Multiclass Users Traffic Flow” 與 9/26 的“The Effect of Driving Behavior on Multiclass Users Traffic Flow”。同時，亦受邀擔任 9/26 日下午 Symposium 17 – “Technology Management”的主持人。此次，參與發表論文的約有四百多篇，內容含括物理、化學、電子、生物、財務、科管等領域之計算應用研究。對於從事計算方面之研究，不論是理論上之推導或實務上之應用均獲得豐富之交流機會。以下為重要行程摘要：

<u>96 年 09 月 25 日</u>

於 Symposium – 16 發表“Classification of Driving Behavior by Pattern Recognition in Multiclass Users Traffic Flow”，此篇論文主要為本年度之成果。於同場所發表之論文亦有利用雷達偵測器辨別車種之研究，在討論中，個別對彼此的研究均提出建議與未來方向，使本研究得以趨於完備。

<u>96 年 09 月 26 日</u>

主持 Symposium 17 主題為 “Technology Management” 發表論文本研究之延伸成果 – 多用路行為對車流之影響，篇名為“The Effect of Driving Behavior on Multiclass Users Traffic Flow”。本論文為利用單胞自動機制模擬多用路車流行為，所用的參數為計畫產生之輸出結果。同一場次亦有同類型之研究，其主要探討之行為在於多車道車流模擬，尤其內容與會後討論，對於未來將研究拓展至多車道多用路行為之車流研究實有助益。

<u>96 年 09 月 27~29 日</u>

參與其他報告場次。

二、與會心得

雖然運輸領域並非本研討會的主要領域，但在參與的會議場次中發覺應該廣泛涉獵各種

理論與應用，尤其是數學與計算，不僅可作為各類型研究之基礎，亦可作為未來研究方向及問題解決的方法。而在電子技術快速發展的今日，電腦的計算能力一日千里，從前窒礙難行的理論，目前大多已可藉助電腦模擬求解。計算與模擬相較於實際的實驗設計可節省大量的時間與金錢，縮短研發或策略研擬時程，為一有效率的研究方法。尤其在交通運輸相關研究，往往涉及大規模的調查，耗費時間、金錢與人力成本，若能有效應用模擬技術，不僅可達到最佳化的設計，亦可節省成本，實為一值得推廣並發展之方法論。