

國立交通大學
運輸科技與管理學系碩士班

碩士論文

最佳權重法應用於交通資料融合

The Optimal Weighting Approach For Traffic

Data Fusion



研究生：吳瑞豐

指導教授：王晉元

中華民國 九十五年 七月

最佳權重法應用於交通資料融合
The Optimal Weighting Approach For Traffic Data Fusion

研究生：吳瑞豐

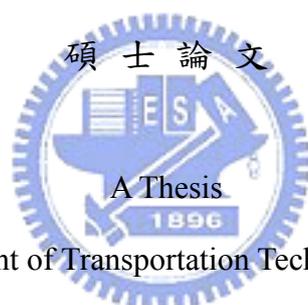
Student : Ruei-Feng Wu

指導教授：王晉元

Advisor : Jin-Yuan Wang

國立交通大學

運輸科技與管理學系



Submitted to Department of Transportation Technology and Management

College of Management

National Chiao Tung University

In Partial Fulfillment of the Requirements

For the Degree of Master of Engineering

In

Transportation Technology and Management

July 2006

Hsinchu, Taiwan, Republic of China

中華民國九十五年七月

最佳權重法應用於交通資料融合

學生：吳瑞豐

指導教授：王晉元

國立交通大學運輸科技與管理學系碩士班

摘 要

先進旅行者資訊系統(ATIS)為智慧型運輸系統(ITS)發展的重點之一，其路況資訊可由許多來源獲得，如偵測器、探針車等，每個來源能提供的資料內容不同，亦有各自的使用範圍及限制，因此不同來源的路況資料必須透過資料融合(Data Fusion)的方法來處理，以得到一個較可靠、較精確的交通資訊。

本研究以最佳權重法來融合多種異質的交通資料，最佳權重法是將系統總不確定性最小化，使得融合後的資料有最低的不確定性，模式中以 Shannon 熵表示不確定性，並提出距離權重法來改善明確分類方式所造成的偏誤。本研究以電腦模擬的方式，來評估模式的適用性，測式結果顯示以距離權重法來分類資料，可降低分類邊界的影響，也可改善資料過度集中或分散所造成的偏誤。

關鍵字：旅行者資訊、資料融合、先進旅行者資訊系統、熵

The Optimal Weighting Approach For Traffic Data Fusion

Student : Ruei-Feng Wu

Advisor : Jin-Yuan Wang

Department of Transportation Technology and Management

National Chiao Tung University

Abstract

Advanced Traveler Information System (ATIS) is one of the key elements of Intelligent Transportation System (ITS). Traveler information could be obtained via various sources such as VD and probe vehicles. Therefore, various information must be merged into a distinct and reliable information. The purpose of this research is developing a data fusion methodology for merging various traffic data source.

The optimal weighting approach is proposed for fusing various traffic data. It minimizes total system uncertainty while fusing data. We use Shannon entropy to represent uncertainty. Besides, a weighted distance approach is used to reduce the impact of data classification.

We use simulation data to evaluate the performance of our model. The results show that our proposed approach could reduce influence of data classification and reduce the bias of centralized and separated data.

Keyword : Traveler Information, Data Fusion, Entropy, ATIS

誌 謝

本研究能夠順利完成，首先要感謝王老師，王老師在學術上的教導自是不在話下，每次都是要讓學生親自試過一遍，才會告訴學生比較好的做法，總覺得從老師身上學到的東西，遠遠超過作研究的範圍，一些處理事物技巧以及待人接物的理道，都是終身受用。

感謝系上的卓訓榮老師以及成功大學的胡大瀛老師，兩位老師在百忙之中撥空來參加我的口試，給予許多寶貴的意見，讓本篇研究能更臻完備。

感謝實驗室裡的學長姊們，從研一開始就給不懂事的我許多幫助，大師兄就像老師一樣很愛上課，小松在計畫案及課業上給予很多的指導，hoho 的超人生活和 CP 值理論對我來說影響深遠，彥佑常在我緊張、沮喪的時候給我鼓勵及支持，還有同梯的威豪、嘉英、小翔、思文，雖然我是實驗室這一屆裡唯一從外系來的，但是對我還是視如己出，在一些不懂的地方給予提醒，也一起完成許多事，還有文誠、浩子哥、詩敏、韻佳，常常在實驗室裡帶給我們歡樂。

最後要感謝我的父母以及兩個哥哥，雖然家裡的經濟狀況不是很好，但是你們還是讓我順利的完成兩年的求學生涯，不用分心擔心家裡的事，這兩年來真是辛苦你們了!!還有郁馥，謝謝你在我忙碌的時候，能夠體諒我，一直陪在我身邊。



吳瑞豐

新竹交大
2006/7/6

目 錄

一、緒論	1
1.1 研究背景與動機.....	1
1.2 研究目的.....	2
1.3 研究範圍.....	2
1.4 研究流程.....	2
二、文獻回顧	4
2.1 資料融合概述.....	4
2.1.1 資料融合的定義與特性.....	4
2.1.2 資料融合的方法與分類.....	5
2.2 資料融合演算法.....	6
2.3 文獻回顧小結.....	10
三、模式構建	11
3.1 最佳權重法(The Optimal Weighting Scheme)	12
3.2 熵(Entropy).....	13
3.3 最佳權重法應用於交通資料融合.....	15
3.3.1 資料分類及熵的計算.....	15
3.3.2 權重的計算.....	19
3.4 資料分類之修正.....	20
四、模擬測試	27
4.1 模擬資料之產生.....	27
4.2 實驗設計.....	31
4.3 測試結果.....	33
4.3.1 情境一.....	33
4.3.2 情境二.....	35
4.3.3 情境三.....	37
4.4 模擬測試小結.....	39
五、結論與建議	40
5.1 結論.....	40
5.2 建議.....	40
六、參考文獻	42

圖目錄

圖 1.1 研究流程圖	3
圖 3.1 資料分佈於分類邊界示意圖	16
圖 3.2 資料平移示意圖	17
圖 3.3 分類偏誤示意圖 1	20
圖 3.4 分類偏誤示意圖 2	21
圖 4.1 常態分配產生流程圖	30



表 目 錄

表 2.1 資料融合方法	6
表 2.2 傳統條件機率理論與TBM之差異	8
表 3.1 道路服務水準分類	15
表 3.2 修正後之分類	15
表 3.3 資料平移範例	18
表 3.4 資料分類範例	18
表 3.5 資料融合範例	19
表 3.6 分類偏誤範例 1	21
表 3.7 分類偏誤範例 2	22
表 3.8 分類偏誤範例 3	23
表 3.9 距離權重法範例	24
表 3.10 修正分類偏誤範例 1	25
表 3.11 修正分類偏誤範例 2	26
表 3.12 修正分類偏誤範例 3	26
表 4.1 GPS車速資料	28
表 4.2 卡方檢定頻率分佈計算	29
表 4.3 實驗設計—情境一	31
表 4.4 實驗設計—情境二	32
表 4.5 實驗設計—情境三	32
表 4.6 模擬車速資料—情境一	33
表 4.7 分類次數—情境一	34
表 4.8 分類機率—情境一	34
表 4.9 熵與權重—情境一	34
表 4.10 模擬車速資料—情境二	35

表 4.11 分類次數—情境二	36
表 4.12 分類機率—情境二	36
表 4.13 熵與權重—情境二	36
表 4.14 模擬車速資料—情境三	37
表 4.15 分類次數—情境三	38
表 4.16 分類機率—情境三	38
表 4.17 熵與權重—情境三	38



一、緒論

1.1 研究背景與動機

即時路況資訊無論在都會區或城際間之交通運輸上，皆扮演著很重要的角色。近年來，車輛的數目隨著經濟的發展而快速成長，使得道路上常常發生交通擁擠的情況，駕駛者也往往會遇到塞車的困境。有了即時路況資訊，不但能使交通管理者掌握現有之交通狀況，做出最佳交通管理與控制策略，也能讓駕駛者得知目前的路況，作為決定出發時間與選擇路線的參考。為了使得交通管理者與駕駛者能充分的了解現有的交通狀況，要如何獲得即時且較精確之即時路況資訊，為一個相當重要的課題。

即時路況資訊可由許多來源獲得，如固定式車輛偵測器、紅外線偵測器、影像偵測器、GPS 探針車等，每種偵測器能提供的資料內容各有不同，亦有各自的使用範圍及限制。然而，在同一時段、同一道路上，不同偵測器所收集到的路況資料往往不盡相同，造成這種現象的原因有很多：第一，收集到的資料樣本不同，例如固定式車輛偵測器能獲得經過偵測器上方的車輛資料，而探針車所收集到的為裝設有 GPS 的車輛資料；第二，偵測器皆存在著誤差，而且每個偵測器的誤差亦不相同，此與雜訊的影響與儀器本身的誤差有關。

由於各個偵測器的即時路況資料往往不盡相同，若將不同的資訊一一發佈給駕駛者，反而容易造成駕駛者的混淆，無法讓駕駛者有效地運用，來決定他們的出發時間及旅行路線。因此不同來源的路況資料必須透過資料融合(Data Fusion)的方法來處理，將各偵測器的資料融合為一個較可靠、較精確的交通資訊，才能便於交通管理者與駕駛者來使用。

資料融合為利用一個有系統的方法，給予各資料來源適當的權重，用以整合不同來源的資料，得到一個較精確、可靠之資訊的一種技術，其中包括資料的收集、過濾、分析以及結合。透過資料融合的處理程序，可以降低資訊的不確定

性，並提供較大範圍的資訊。資料融合的技術起初是用於軍事上的用途，後來便逐漸應用到各個領域，其中也包括交通資料融合上的應用，過去曾有許多的研究對資料融合的技術做探討，但各有使用上的限制及誤差。因此本研究期望能發展一資料融合模式，用以整合交通資料，進而提供較可靠的即時路況資訊。

1.2 研究目的

本研究的主要目的在於發展一套合適的資料融合模式，用以整合不同來源的交通資料(如：固定式車輛偵測器、影像偵測器、紅外線偵測器、GPS 探針車等)，以提供駕駛者較精確、可靠之路況資訊。

1.3 研究範圍

資料融合的流程包括有資料的收集、過濾、分析以及結合，本研究僅考慮結合的部分，對於資料收集、過濾及分析方面則不加探討。另外，交通資料的種類很多，包括時間平均速度(Time Mean Speeds)、車間距(Headway)、車輛種類、車道佔有率(Lane Occupancy)、車輛位置，車輛速度等，本研究發展的資料融合模式可應用於各種交通資料融合上，在之後的章節則以平均車速為例來做探討。

1.4 研究流程

本研究之研究流程可分為五個部分：問題定義、文獻回顧、模式建立、模式驗證以及結論與建議，流程圖如圖 1.1 所示。

首先，我們要對欲研究的課題做一個完整的敘述，以了解研究的目的及範圍，之後針對問題的特性做相關的文獻回顧，藉由參考其他學者的研究成果，來發展一套合適的交通資料融合模式，然後使用電腦程式來產生模擬資料，對模式做細部的修正，並利用模擬資料來評估模式的成效，並對研究結果做結論與建議。

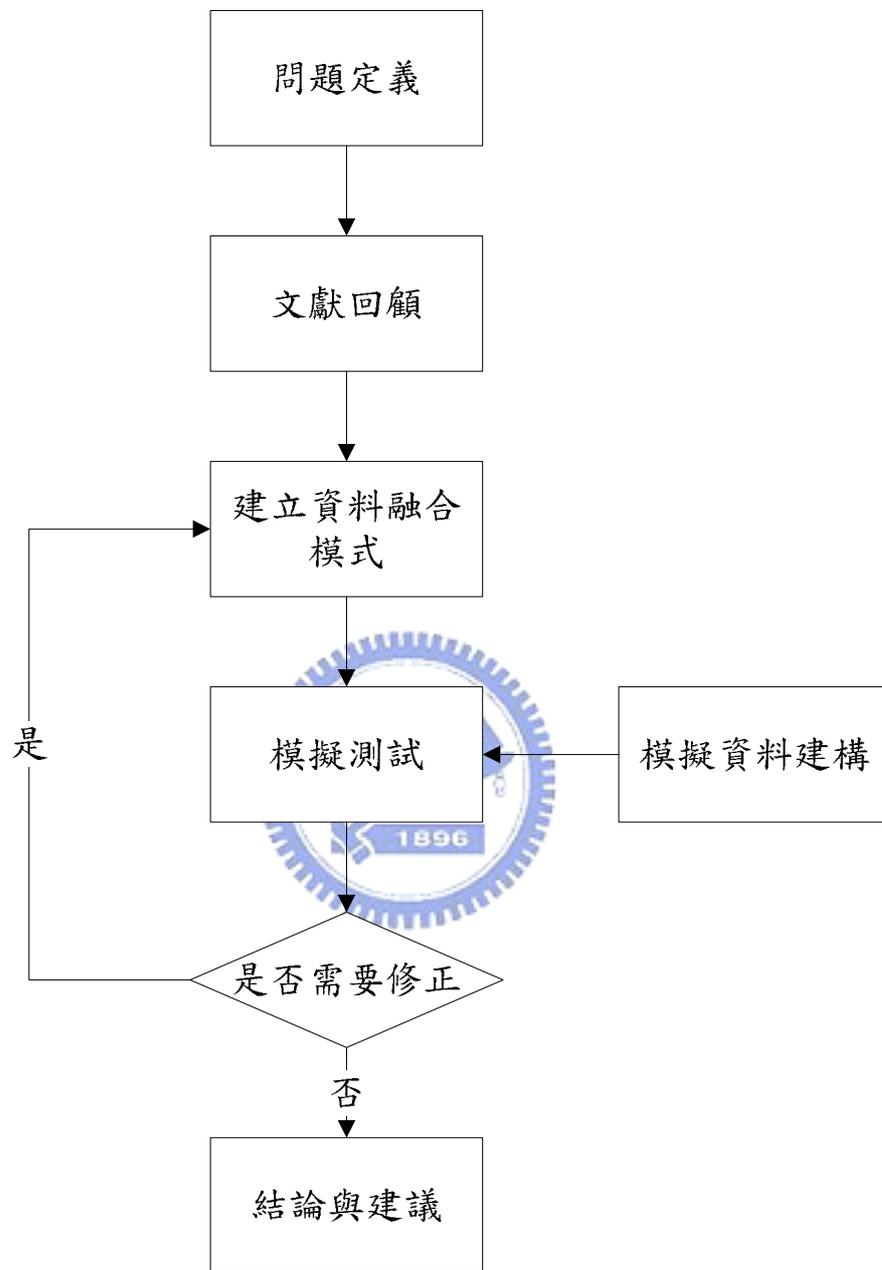


圖 1.1 研究流程圖

二、文獻回顧

本研究主要在以發展適合的資料融合模式，用來融合數個不同來源回傳的交通資料，以推估路段的行駛速率，提供駕駛者作為旅行時間的參考。因此，本章將針對資料融合之相關文獻進行探討，在 2.1 節對資料融合做一個概述，2.2 節則著重於資料融合演算法的介紹。

2.1 資料融合概述

資料融合的發展開始於 1980 年代後期，美國國防部最初將資料融合用於軍事偵查與基地作戰管理系統[8]。近年來，陸陸續續有研究將其應用於不同的領域，如機器人、影像處理及其他非軍事用途，其中也包括資料融合在交通管理與智慧型運輸系統上的應用。

2.1.1 資料融合的定義與特性



美國國防部(U. S. Department of Defense)對「資料融合」的定義為一種有關自動偵測、處理、相關、估計及單一或是多種資料來源之整合，進行多層級、多事件之處理程序[8]。在美國華盛頓州交通部(Washington State Department of Transportation)的 ITS 資料融合報告書中，則將「資料融合」定義為利用中央資料處理器來整合數個來源的資料，以提供綜合且精確的資訊[3]。

根據 Sarma and Raju 的研究[12]，資料融合具有下列特性：

1. 提高資訊的信賴程度：以兩個以上的偵測器來確認同一個目標，可提高對目標特性描述的信賴程度。
2. 降低不確定性：數個偵測器的聯合資訊能減少目標特性推論的可能情況。
3. 改善偵測效果：整合多個觀測值能改善訊號與雜訊的比例，增加偵測的正確性。
4. 增加實用性：當有部分偵測器失效時，還是可以藉由其他可用之偵測器來提

供資訊。

5. 增加時間與空間的範圍：各種偵測器的可用範圍不同，利用不同範圍的偵測器來互相彌補不足的地方，提供較大範圍的資訊。
6. 降低成本：多個偵測器的平均成本會遠低於一個相同精確度的偵測器。

2.1.2 資料融合的方法與分類

資料融合可用單一或是多種處理程序來整合資料，也依提供給使用者資訊的程度不同而有所區分。美國國防部將資料融合程度分為三個層級[9]：

- 層級一：對收集到的原始資料做篩選及過濾，將極端值與不符合常理的部分刪除或加以修正，使得修正後的資料較能代表實際的情況。以路段交通資料來說，收集到的原始資料可能包括車輛因紅綠燈減速或停等的部分，以及偵測器誤差所造成的異常值，所以在使用這些資料的時候，必須先加以修正才較能符合實際的路段資訊。
- 層級二：將兩個或兩個以上來源的資料，給予適當的權重，使其融合以提供一個較可靠的資訊。固定式車輛偵測器與 GPS 探針車為國內最常見的路段交通資訊來源，但兩者的特性、使用範圍、誤差及同時段內的資料數皆不盡相同，若將兩者收集到的資料以相同的權重融合，來提供路段交通資訊，為一較不合理的做法，因此，使用層級二的資料融合方法來決定各資料來源的權重再加以融合，對路段交通資訊較能做精確的推估。
- 層級三：對層級二的資訊加以解釋及說明，並提供解決的方案供使用者參考。在路段交通資料上，層級二的資料融合能提供駕駛者路段上的平均行駛速率，而層級三的資料融合能更進一步地告訴駕駛者此路段是否擁塞，並於擁塞時建議駕駛者改走其他較順暢之道路，以節省旅行時間。

根據上述三個層級，Linn and Hall 將資料融合的演算法分為五大類[9]：資料關聯、狀態估計、認證融合、形態辨識及人工智慧，其中較常見的資料融合方法及分類如表 2.1 所示。

表 2.1 資料融合方法

融合層級	融合方法	融合技術
層級一	資料關聯	Figure of merit Gating techniques
	狀態估計	Kalman filters
層級二	認證融合	Bayesian decision theory Dempster-Shafer evidential reasoning Adaptive neural networks
	形態辨識	Cluster methods
層級三	人工智慧	Expert systems Blackboard architecture Fuzzy logic

資料融合方法有很多，而要選擇何種方法來整合資料，則需視收集到的資料特性，以及欲提供給使用者的資訊層級。本研究之主要目的在於融合不同來源的交通資料，以提供較正確的資訊給道路使用者，因此層級二的資料融合方法，較適合本研究使用，之後在資料融合演算法的探討上，則著重於層級二的資料融合方法。

2.2 資料融合演算法

Huadong et al.[5]於 2002 年提出以權重 Dempster-Shafer 理論(Weighted Dempster-Shafer Theory)，利用會議中聲音及影像，判斷正在講話的人與欲交談的對象。Dempster-Shafer 理論是一種以統計為基礎之資料融合與分類演算法，主要原理在於當偵測器或其他交通資料來源無法確定其事件之機率為百分之百時，在演算法的過程中，計算各偵測器的事件分類機率，再依 Dempster 法則進行事件間相互連結之關係。Huadong et al.比較了線性加總、DS 理論以及權重 DS

理論，結果發現三者判斷的正確性皆比單以影像或聲音來判斷的情況好，但這三種方法卻無顯著的差異。

2003 年 Huadong et al.[6]更進一步地提出動態權重 Dempster-Shafer 理論 (Dynamic Weighted Dempster-Shafer Theory)，以同樣的實驗數據來比較此四種方法(線性加總、DS 理論、權重 DS 理論及動態權重 DS 理論)的成果，結果顯示，以動態權重 DS 理論來融合資料，有較高的正確率。

2004 年曾治維[15]將 Dempster-Shafer 理論應用於交通資料整合上，為了解決兩資料來源的衝突，此模式先將兩組資料平移，依照其樣本數與標準差的特性，移至同一平均值之後，再處理資料融合的程序。此研究結果發現雖然其模式能降低兩資料來源的衝突，但卻容易將資料平移至樣本數較多之來源的平均值，若在其誤差較大的情況下，融合後的結果則較不盡理想。

Andrej et al.[1]在 1999 年提出以 TBM(Transferable Belief Model)來找出生產線上的錯誤，並將其應用於直流馬達的錯誤測試。在銷售市場上，如何提高產品的品質以及降低產品的製造成本是相當重要的，在一個自動化生產工廠裡，一條龐大生產線也許有幾千幾萬個組件，我們無法精確的預測錯誤可能發生的地方，TBM 就是在處理這種未知假設的問題。TBM 的概念與傳統的條件機率理論相似，兩者的差別在於，傳統的條件機率理論會先定義出可能出現的集合，並且需要各集合的先驗機率(priori probability)，而 TBM 會給予未知的集合一個機率函數，且不需要各集合的先驗機率。以表 2.2 來說，假設有兩個偵測器的推論結果不相同，一個偏向於假設一，一個偏向於假設三，但是依傳統條件機率的融合方式，卻會認同假設二，如此就有可能會造成偏誤，而 TBM 則會給予未知的集合一個機率函數。研究的結果認為 TBM 容易與模糊邏輯結合應用，且測試的結果發現，以 TBM 結合模糊邏輯的方法來診斷系統錯誤，正確率優於布林推論 (Boolean Reasoning)的結果。

表 2.2 傳統條件機率理論與 TBM 之差異

	推論一	推論二	傳統條件機率	TBM
假設一	0.99	0	0	0
假設二	0.01	0.01	1.00	0.0001
假設三	0	0.99	0	0
未知的假設	---	---	0	0.9999

2001 年 Ruey et al.[11]用類神經網路法(Neural Networks)融合 GPS 探針車與固定式車輛偵測器的資料，推估主要幹道上的平均車速。類神經網路是一種計算系統，它使用大量簡單的相連人工神經元來模仿生物神經網路的能力。人工神經元是生物神經元的簡單模擬，它從外界環境或者其他人工神經元取得資訊，加以簡單的運算，並輸出其結果到外在環境或其他神經元。它是由相當簡易的基本元件組成，各基本元件均能進行資訊之接收、處理、傳遞、儲存等功用。在 Ruey et al.的測試範例中，使用了 1032 筆的歷史資料來校估權重，其融合後的結果與單一來源估計相比，可降低超過 50%的誤差，但其模式無法修正車速因車輛停等紅綠燈所造成的誤差。

2002 年 Keechoo and YounShik[7]利用 GPS 探針車與固定式車輛偵測器的歷史資料，以模糊回歸(Fuzzy Regression)的方式取兩資料來源的權重，來估計路段上的平均車速。由於探針車與車輛偵測器回傳資料的頻率不相同(其測試範例之探針車回傳頻率為五分鐘，而車輛偵測器的回傳頻率為一分鐘)，因此 Keechoo and YounShik 在融合資料前，必須先對車輛偵測器的資料做處理，用投票法(Voting Technique)將其修正為五分鐘的資料，再進行融合，其公式如下：

$$w_k = \frac{n_k}{s_k^2}$$

其中 w_k 為每一分鐘資料的權重， n_k 為每一分鐘的資料數， s_k^2 為每一分鐘資料的標準差。

之後再用模糊回歸的方式取兩資料來源的權重，來估計路段上的平均車速。Keechoo and YounShik 用 MAPE(Mean Absolute Percentage Error)的方式來比較模糊回歸與幾何平均的測試結果，認為單純以幾何平均的方法來推估路段車速，無法考慮車輛左轉以及公車上下客人的停等，而模糊回歸可將這部分的車速，給予較低的歸屬函數，來反應因上述情形所造成的延滯。

1992 年 Otman and Helen[10]為了解決這個問題，提出以團隊共識法(Team Consensus Approach)來降低機器人內各偵測器觀測值的衝突。資料融合的技術在機器人的領域上應用的相當廣泛，為了能使機器人清楚的了解周遭的情況，機器人身上通常都會有許多的偵測器來量測附近物體的物理特性(如：位置、速度、加速度等)，而在這些偵測器的量測結果中，時常會發生衝突，或者是單一偵測器的準確度不高的情況，因此如何降低機器人內各偵測器觀測值的衝突，並提高觀測值的正確性，就變成相當重要的一個課題。團隊共識法主要的概念是以可能出現的集合機率來表示效用函數，由於每一個效用函數皆會影響其他偵測器的效用函數，因此必須對其他的效用函數加以修正，直到收斂為止，如此各個資料來源則會得到一個最佳的權重。Otman and Helen 用 Shannon 熵(Shannon Entropy)來求出修正的權重。熵為信息理論 (Information Theory)的重要觀念，藉以描述關於不確定性的數學度量，Otman and Helen 的方法會使得融合後的資料具有最小的熵值，也就是說能降低資訊的不確定性，並以機器人的辨識系統來做模擬測試。

2004 年吳欣潔[14]將熵應用於交通資料的融合，以最佳權重法(The Optimal Weighting Scheme)使得融合後的交通資訊有最小的不確定性，並以道路服務等級對交通資料做分類，計算出各資料來源的熵再加以處理，此研究認為路段上車速觀測值的異常可以反應在不確定性上，但是卻無法得知其固定誤差，且計算的熵值與資料分類的標準有關。

2.3 文獻回顧小結

綜合上述文獻回顧，本研究認為 TBM 雖然能夠降低傳統條件機率的偏誤，但是對未知集合的處理，尚未有一套完整的方法，然而交通車速資料可能出現的情況是可以預測，例如在一般道路上，車速通常會在 0~80 公里之間，超過 80 公里以上的觀測值，可視為異常值而刪除，因此不需處理未知集合的問題。

Dempster-Shafer 理論來融合交通資料，較無法處理各來源之衝突，若兩資料來源的平均值差異甚大，很容易會認同發生機率較小的集合，而造成偏誤，曾治維雖然在融合前先將一組資料做平移，以降低資料來源的衝突，但是在未知哪一個偵測器較精確的情況下，而做資料平移的處理，也有發生偏誤的可能

類神經網路及模糊回歸的方法則需要大量的歷史資料來校估參數，而交通實際路況資料的收集實屬不易，一般的做法是利用人工於道路上以測速槍收集資料，或者是用攝影的方式，以影像處理來獲得實際的車速，之後再與歷史資料配合以校估參數，然而此兩項做法皆需耗費大量的人力、物力來取得。

最佳權重法能將異常的觀測值表現在熵值上，且不會受到資料平均值的影響，唯其亦受到資料分類的限制。因此本研究期望修正最佳權重法之交通資料分類方式，發展出一套較合適的資料融合模式。

三、模式構建

本章節將提出整合不同偵測器之交通資料融合模式，並以路段平均車速為例來做探討。

從各偵測器取得的交通資料中，由於取樣差異與儀器誤差的不同，所以各來源的資料特性(如樣本數、平均值、標準差等)亦有所不同，若我們將所有資料來源的平均車速直接取平均值，來代表此時之路段平均車速，則有失其合理性，因此交通資料融合常以加權平均的方式來處理。加權平均法為給予各偵測器適當的權重再進行加總，如 3.1 式，首先我們可以從偵測器 i 收集的資料中，得到路段平均車速 V_i ，再依各偵測器收集到的資料特性，給予不同的權重 W_i ，予以加總，求得融合後的平均車速 V 。

$$V = \sum_{i=1}^N W_i V_i \quad (3.1)$$

由 3.1 式可知，融合後的平均車速會受到給予的權重所影響，因此要如何決定各偵測器適合的權重 W_i ，為資料融合之重點所在。

本研究以最佳權重法來構建交通資料融合模式，計算每個資料來源的權重，將不同的資料加以融合。最佳權重法的原理在於將系統的不確定性最小化，使得融合後的系統有最低的不確定性。因此在用最佳權重法來獲得各偵測器的權重前，必須先量測各偵測器的不確定性。

文獻中以 Shannon 熵來代表各偵測器的不確定性，應用於平均車速資料融合上[14]，Shannon 熵為信息理論(Information Theory)中不確定性的量測方法，利用樣本空間集合發生的機率，來計算每組資料的不確定度。

綜合上述可知，本研究的資料融合模式主要分為熵的計算以及最佳權重法兩個部分。本章 3.1、3.2 節先對最佳權重法及熵的基本概念做一個簡單的敘述，3.3 節介紹吳欣潔如何以最佳權重法將熵值最小化，應用於交通資料的融合上

[14]，3.4 節則對上述方法的資料分類方式提出修正，以降低融合後資訊的偏誤。

3.1 最佳權重法(The Optimal Weighting Scheme)

最佳權重法的原理在於利用給予各資料適合的權重，將系統的總不確定性最小化，使得融合後資訊的不確定性達到最低。假設每個偵測器 i 的不確定性為 H_i ，則整個系統的不確定性為

$$T = \sum_{i \in S} H_i \quad (3.2)$$

為了給予每個偵測器適當的權重 W_i ，文獻中以最小平方法使得系統中所有偵測器的不確定性平方和為最小[10]，其最佳化問題如下：

$$\text{Minimize} \quad \sum_{i \in S} W_i^2 H_i^2 \quad (3.3)$$

$$\text{Subject to} \quad \sum_{i=1}^N W_i = 1 \quad (3.4)$$

$$W_i > 0 \quad (3.5)$$

3.3 式將 $\sum_{i \in S} W_i^2 H_i^2$ 最小化，會使得不確定性較高的偵測器，對應到較小的權重。3.4、3.5 式則表示所有偵測器權重的總和為 1，且各個偵測器的權重皆大於 0。

此最佳化問題可經拉氏鬆弛法(Lagrangian Relaxation)推得：

$$Z = \sum_{i \in S} W_i^2 H_i^2 - \rho [\sum_{i \in S} W_i - 1] \quad (3.6)$$

將 Z 分別對 W_i 與 ρ 偏微分，使一階導函數為 0，求得：

$$W_i = \frac{\rho}{2H_i^2} \quad (3.7)$$

$$\sum_{i \in S} W_i(\gamma) = 1 \quad (3.8)$$

綜合上列 3.7、3.8 式：

$$\sum_{i \in S} \frac{\rho}{2H_i^2} = 1 \quad (3.9)$$

$$\Rightarrow \rho = \frac{2}{\sum_{i \in S} H_i^{-2}} \quad (3.10)$$

將 3.10 式代入 3.7 式得：

$$W_i = \frac{1}{H_i^2 \sum_{i \in S} H_i^{-2}} \quad (3.11)$$

3.11 式即為各偵測器之最佳權重。由 3.11 式的結果可以發現，各偵測器的權重與不確定性的平方和成反比，因此不確定性高的資料，會分配到較低的權重；反之，不確定性低的資料，則會分配到較高的權重，利用此方法來降低總系統的不確定性。由 3.11 式可知，權重 W_i 為不確定性 H_i 的函數，因此只要能求得資料的不確定性 H_i ，即可求得權重 W_i 。

3.2 熵(Entropy)

熵為一種不確定性的表示方式，最早為 Clausius 提出，當時是用來量測熱力系統中的能量。到了 1940 年代後期，由於信息理論的需要而首次出現了 Shannon 熵，藉以描述關於不確定性的數學度量。

Shannon 假設樣本空間 X 中有 n 個事件，其中每一個事件 w_i 發生的機率為 p_i ， $i=1,2,\dots,n$ ，為了量測此樣本空間的「不確定度」，則定義一個函數

$H(p_1, p_2, \dots, p_n)$ ，其定義域是所有的樣本空間。若要精確地反映試驗結果的不確定度，則 H 必須滿足下列三個基本條件：

1. 對固定 n 來說， H 是 p_1, p_2, \dots, p_n 的連續函數。
2. 若 $p_i = \frac{1}{n}$ ， $i = 1, 2, \dots, n$ ，則對應的 $H\left(\frac{1}{n}, \dots, \frac{1}{n}\right)$ 應該為 n 的單調遞增函數。
3. 若某一試驗分解成多個試驗，則原先的 H 值應為相應的各個 H 值之加權和。

滿足條件 1、2 和 3 的函數 $H(p_1, p_2, \dots, p_n)$ 恰好為 $-k \sum_{i=1}^n p_i \log_c p_i$ 的形式，其中 K 為正的常數。Shannon 則以 $k = 1$ 、 $c = 2$ 來做為其熵值：

$$H(p_1, p_2, \dots, p_n) = -\sum_{i=1}^n p_i \log_2 p_i \quad (3.12)$$

Shannon 熵具有下列幾項特性：

1. 當一個機率為 0 的集合增加時，Shannon 熵不會改變。

$$H(p_1, p_2, \dots, p_n) = H(p_1, p_2, \dots, p_n, 0)$$

2. 若各集合的機率分佈的排列改變，Shannon 熵不會改變。

$$H(p_1, p_2, \dots, p_n) = H(\text{perm}(p_1, p_2, \dots, p_n))$$

3. Shannon 熵為一連續的函數。
4. 若可能發生事件的數目固定，則最大的熵值出現在均勻分配(uniform distribution)的時候。

$$H(p_1, p_2, \dots, p_n) \leq H\left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}\right)$$

5. 若所有集合的機率相同，則可能出現的集合數愈多，Shannon 熵愈大。
6. Shannon 熵為非負數，且為一凹形函數。
7. 若某一事件的發生機率為 1，則熵值為 0。

3.3 最佳權重法應用於交通資料融合

3.3.1 資料分類及熵的計算

由於本研究之模式是利用各資料之熵值來決定融合的權重，因此我們必須先計算各偵測器的熵，才能進一步地來推估權重。由 3.12 式可知，計算熵值必須要有樣本空間內各集合發生的機率，然而路段平均車速為一連續性分佈的資料，因此必須先對蒐集到之原始交通車速資料分類，才能進一步做熵值的計算，再依各資料來源的熵來獲得權重。

根據交通部公路容量手冊，交通資料可依不同的道路等級，將服務水準分成六級，如表 3.1 所示，每個等級之服務水準各有不同的標準。為了簡化模式，吳欣潔則將資料分類修正為三類[14]，如表 3.2 所示。

表 3.1 道路服務水準分類

道路等級	I	II	III
自由車流速率(kph)	55	45	40
道路服務水準	平均車速(kph)		
A	>51	>43	>33
B	39~51	32~43	25~33
C	34~39	27~32	20~25
D	29~34	23~27	16~20
E	21~29	17~23	10~16
F	<21	<17	10

表 3.2 修正後之分類

道路等級	I	II	III
自由車流速率(kph)	55	45	40
道路服務水準	平均車速(kph)		
A	>40	>30	>25
B	30~40	20~30	16~25
C	<30	<20	<16

然而，偵測器收集到的交通資料，有時候會有平均車速落在分類邊界附近的情況，如圖 3.1 所示。如此一來，在資料分類的時候容易將相近的車速資料，分在不同的分類裡，由此分類計算出來的熵，儘管此時段資料分佈的變異性小，還是有可能會得到相當高的不確定性，因此為了降低資料分佈位置對融合結果的影響，必須先對收集到的車速資料做平移的調整，調整方法如下：

1. 計算各偵測器所得資料之平均車速。
2. 找出道路分類範圍之中心值。
3. 計算平均車速與中心值的差值 D ，然後將每一筆資料平移 D ，使平移後之平均車速恰為道路分類範圍之中心值，如圖 3.2。

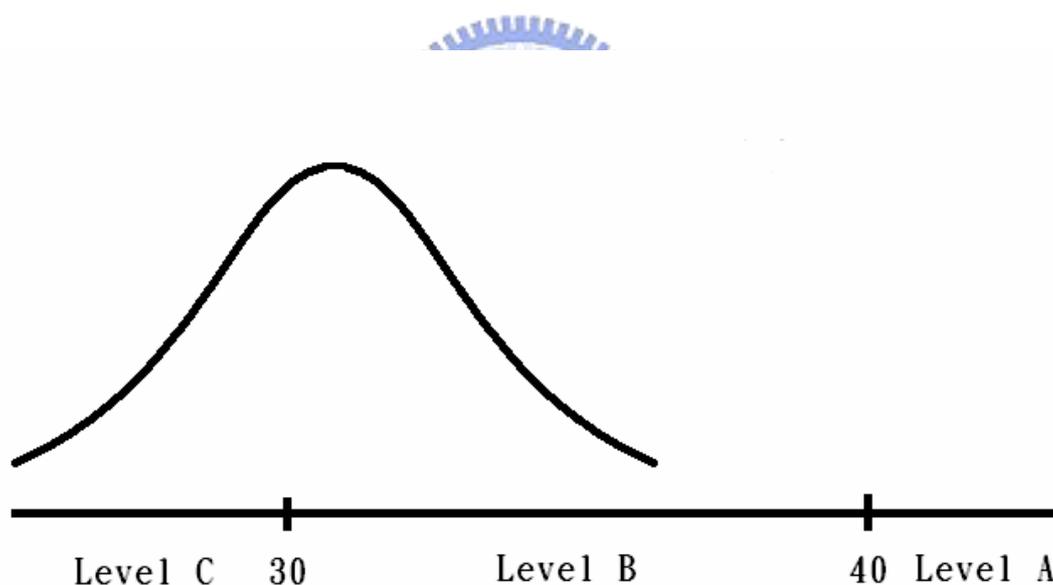


圖 3.1 資料分佈於分類邊界示意圖

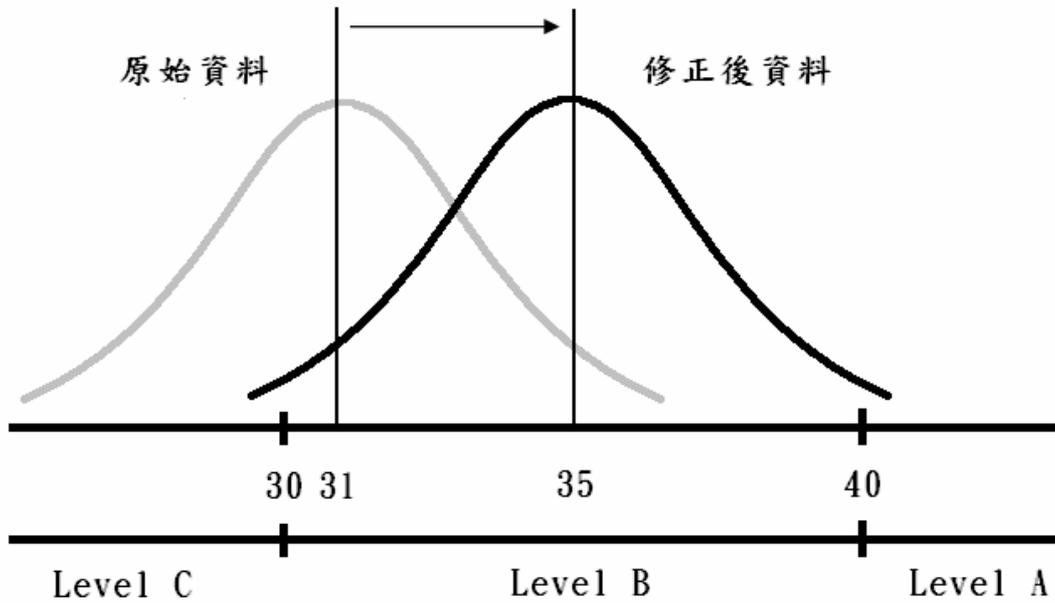


圖 3.2 資料平移示意圖

經過上述之資料平移，就能降低熵值受到資料分佈位置的影響，而僅以資料的變異程度來決定資料的不確定性。表 3.3 為資料平移的範例，我們採用道路等級 I 的分類來做示範，步驟如下：

1. 計算原始資料的平均車速：30.86kph
2. 道路分類範圍之中心值：35kph
3. 計算平均車速與中心值的差值 D ： $35 - 30.86 = 4.14$ kph，再將每筆資料均加上 4.14 kph，使得經平移後的平均車速為分類範圍中心值 35kph。

表 3.3 資料平移範例

原始資料		修正後資料	
24.64	30.43	28.78	34.57
37.45	36.61	41.59	40.75
26.88	34.41	31.02	38.55
37.94	32.34	42.08	36.48
24.12	33.39	28.26	37.53
26.92	29.57	31.06	33.71
36.93	37.38	41.07	41.52
17.99	30.80	22.13	34.94
24.17	35.35	28.31	39.49
26.97	21.53	31.11	25.67
36.40	39.33	40.54	43.47
25.96	28.76	30.10	32.90
37.13	33.32	41.27	37.46
24.94	36.00	29.08	40.14
34.37	23.81	38.51	27.95
30.86		35.00	

在資料平移之後，即對修正後的交通資料依分類來做統計，以求得各分類之機率函數，方法如下：

1. 對此時段內修正後之交通車速資料作次數統計，計算出此時段的資料出現於各分類中的次數。
2. 依各分類中的次數，計算此時段之車速位於各分類之機率函數。表 3.4 為表 3.3 之資料統計範例。

表 3.4 資料分類範例

分類	次數	機率
A	9	0.300
B	14	0.467
C	7	0.233
總計	30	1.000

有了各分類的機率函數之後，就可以來計算此組資料的熵。熵是被用來描述資訊的不確定性，Shannon 熵定義為

$$H(p_1, \dots, p_n) = -\sum_{i=1}^n p_i \log_2 p_i$$

其中 $H(p_1, \dots, p_n)$ 為熵， p_i 為每個可能出現集合的機率。

依照 Shannon 熵的計算方法，可計算出表 3.4 之資料的熵值為 1.524，即為此組資料的不確定性之度量。

3.3.2 權重的計算

在求得熵值後，我們再進一步的利用熵值來計算各資料來源所得之平均車速的權重，並用此權重推估融合後之路段平均車速。由 3.11 式可知，最佳權重法求得各偵測器之最佳權重為

$$W_i = \frac{1}{H_i^2 \sum_{i \in S} H_i^{-2}}$$

求得各偵測器之最佳權重後，再依 3.1 式來融合各偵測器的平均車速，得到融合後的平均車速。表 3.5 為資料融合之範例，由範例中可以發現，經由最佳權重法的計算，熵值較大的偵測器(Sensor 1)則會得到較低的權重。

表 3.5 資料融合範例

資料來源	平均速率	熵	權重	融合後之平均速率
Sensor 1	30.86	1.524	0.374	33.27
Sensor 2	34.71	1.179	0.626	

3.4 資料分類之修正

雖然上述之資料融合模式在計算熵值前，先將原料資料平移，降低了車速資料分佈位置的影響，然而在某些情況時，此種分類方式對於熵及權重的計算，還是容易造成偏誤。

由於交通資料為一連續性的分佈，當資料位於分類邊界附近的時候，若將此筆資料明確的分於某一類，則有失偏頗，且交通資料能收集到的樣本數通常不多，因此每一筆交通資料對於機率函數的敏感度很大[14]，因此當車速資料集中，且有部分資料位於分類邊界附近的時候，即使兩組資料的變異程度相差不遠，還是有可能獲得相差很大的權重，如圖 3.3、表 3.6 所示。在此偏誤範例中，兩偵測器資料的標準差分別為 3.07 與 2.88，由於標準差的差異不大，理論上兩者融合的權重應該不會相去甚遠，但是兩者的資料皆集中於分類 B，而且在樣本數不多的情況下，雖然兩者各分類的資料數相差甚少，卻對熵值影響很大，因此計算出來的權重分別為 0.225、0.775，如此變異程度相差不大的兩組資料，其中一組的權重卻近八成，實為一不合理之現象。

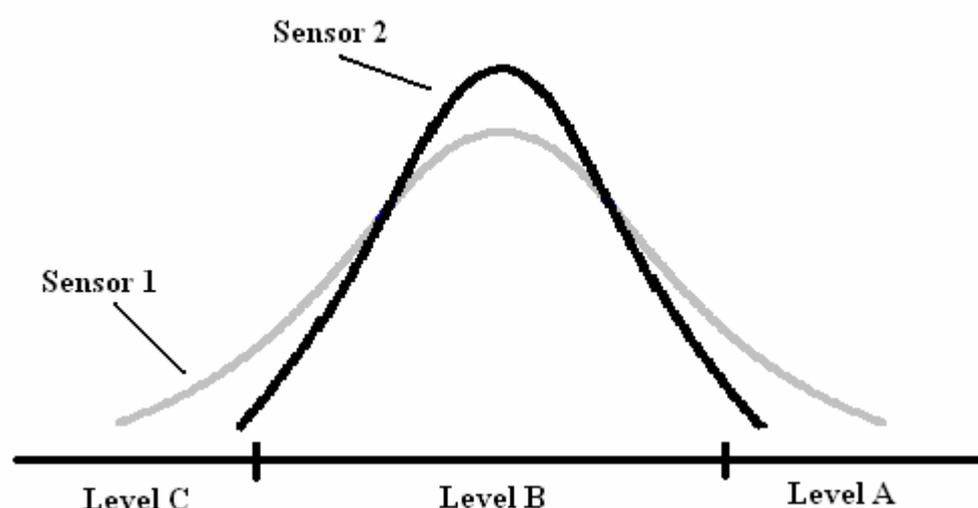


圖 3.3 分類偏誤示意圖 1

表 3.6 分類偏誤範例 1

資料來源	Sensor 1		Sensor 2	
標準差	3.07		2.88	
	次數	機率	次數	機率
A	1	0.033	0	0.000
B	26	0.867	28	0.933
C	3	0.100	2	0.067
熵	0.675		0.363	
權重	0.225		0.775	

或者當某一組資料的車速皆屬於同一個集合(其中一個集合的機率函數為 1)時，則權重的計算則會變成無意義(熵值為 0)，如圖 3.3、表 3.7 所示，在此偏誤範中，Sensor 2 的資料完全落於 B 級中，所以 Sensor 2 的不確定性為 0，權重為無意義，如此便無法做資料融合的处理。

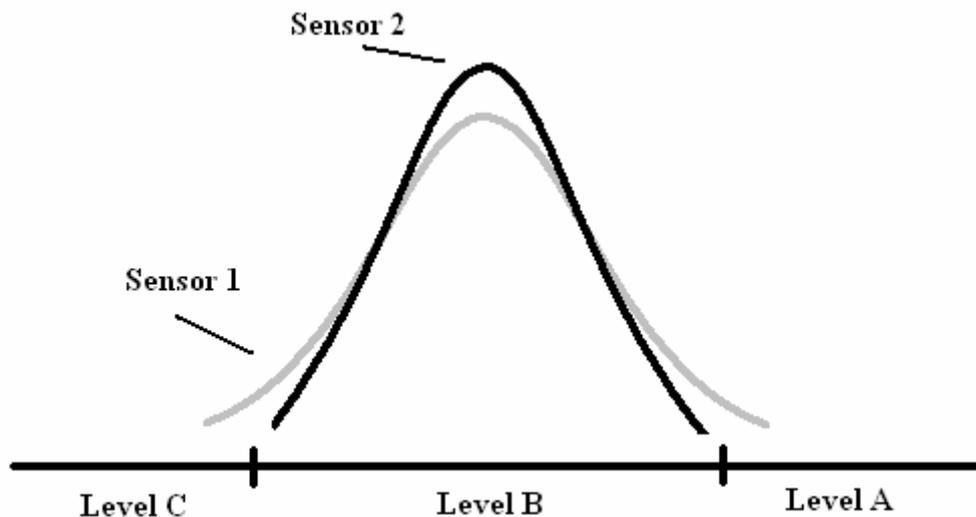


圖 3.4 分類偏誤示意圖 2

表 3.7 分類偏誤範例 2

資料來源	Sensor 1		Sensor 2	
標準差	5.18		1.81	
	次數	機率	次數	機率
A	4	0.133	0	0.000
B	23	0.767	30	1.000
C	3	0.100	0	0.000
熵	1.014		0.000	
權重	---		---	

另外，當資料的變異程度過大時，由於 B 類的分組區間較小，平移後 A、C 分類的資料筆數會大於 B 類，此時由於 A、C 類的機率函數提高，容易使得熵下降，會造成變異程度大，但權重卻較高的情況，如圖 3.5、表 3.8 所示。在此偏誤範中，Sensor 2 的標準差為 15.12，較 Sensor 1 的 9.41 高，但由於 Sensor 2 的資料變異程度過大，使得 A、C 分類的機率函數提高，最後融合的權重反而較 Sensor 1 大。

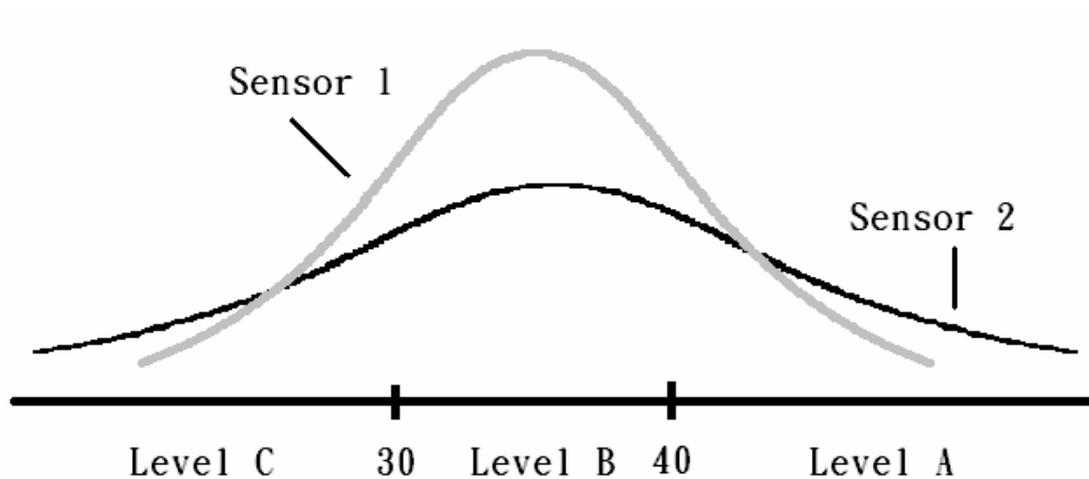


圖 3.5 分類偏誤示意圖 3

表 3.8 分類偏誤範例 3

資料來源	Sensor 1		Sensor 2	
標準差	9.41		15.12	
	次數	機率	次數	機率
A	10	0.333	11	0.367
B	9	0.300	5	0.167
C	11	0.367	14	0.467
熵	1.580		1.475	
權重	0.466		0.534	

由於依文獻中之分類方法會有造成偏誤的可能，因此本研究提出距離權重法來改善上述資料分類所造成的偏誤情況。明確分類法是以 0 或 1 的方式來做資料次數的計算，而距離權重法是利用每一筆車速與相鄰兩類組中點的距離來計算權重，以小數的方式來統計次數。首先，假設有 n 筆車速資料，第 i 筆資料的車速為 V_i ，分配到第 j 類的權重為 N_{ij} ，各分類的組中點為 m_j ，對於第 i 筆資料 V_i 來說，若 V_i 小於第 1 類的組中點 m_1 ，則此筆資料在第 1 類的次數為 1；若 V_i 大於第 n 類的組中點 m_n ，則此筆資料在第 n 類的次數為 1；若 V_i 恰等於某一類的組中點，則此筆資料在該類的次數為 1；若 V_i 在某兩類的組中點之間，則依此車速與這兩類的組中點的距離倒數 D_j^{-1} 、 D_{j+1}^{-1} ，來計算此筆資料分配至這兩類的權重，如下列步驟所示。

令起始值 $N_{ij} = 0$

若 $V_i < m_1$ ，則 $N_{i1} = 1$ 。

若 $V_i > m_n$ ，則 $N_{in} = 1$ 。

若 $V_i = m_j$ ，則 $N_{ij} = 1$ 。

若 $m_j < V_i < m_{j+1}$ ，則 $N_{ij} = \frac{D_j^{-1}}{D_j^{-1} + D_{j+1}^{-1}}$ ， $N_{ij+1} = \frac{D_{j+1}^{-1}}{D_j^{-1} + D_{j+1}^{-1}}$

以一筆車速為 32.8 kph 的資料為例，如表 3.9，其位於 B、C 組中點之間，且與 B、C 組中點的距離分別為 2.2、17.8，依距離權重法可計算出分配至 B、C

組的權重為 $N_B = \frac{1}{2.2(\frac{1}{17.8} + \frac{1}{2.2})} = 0.89$ 、 $N_C = \frac{1}{17.8(\frac{1}{17.8} + \frac{1}{2.2})} = 0.11$ ，且分配之

次數總和為 1。



表 3.9 距離權重法範例

	A (55kph)	B (35kph)	C (15kph)
與組中點之距離	22.2	2.2	17.8
分配各組之權重	0.00	0.89	0.11

以上述之方法，將表 3.6、3.7、3.8 三例經修正後，如表 3.10、3.11、3.12 所示。由表 3.10 可知，兩組標準差為 3.07 與 2.88 的資料，以明確分類法計算出來的權重分別為 0.225 與 0.775，結果過度信賴偵測器二，用距離權重法修正後，求得的權重分別為 0.494 與 0.506，不致於過度信賴某個偵測器，且可將不確定性反應在權重上。另外，觀察表 3.11 的結果發現，原本以明確分類法計算出偵測器二的不確定性為 0，無法求得其權重，而以距離權重法修正後，則可計算出權重。表 3.12 的結果顯示，兩組標準差為 9.41 與 15.12 的資料，以明確分類法計算出來的權重，會使得變異程度大的偵測器獲得較大的權重，用距離權重法修

正後，其權重為 0.559 與 0.441，不致於造成偏誤。

因此用距離權重法來修正各資料機率函數的計算方式，不但可以減少資料分佈大多於分類邊界附近以及變異程度過大的偏誤，也可以計算當 B 組的機率為 1 時的權重，改善上述之情況。

表 3.10 修正分類偏誤範例 1

	資料來源	Sensor 1		Sensor 2	
	標準差	3.07		2.88	
明確 分類法		次數	機率	次數	機率
	A	1	0.033	0	0.000
	B	26	0.867	28	0.933
	C	3	0.100	2	0.067
	熵	0.675		0.363	
	權重	0.225		0.775	
距離 權重法		次數	機率	次數	機率
	A	1.79	0.060	1.76	0.059
	B	26.42	0.881	26.48	0.883
	C	1.79	0.060	1.76	0.059
	熵	0.647		0.640	
	權重	0.494		0.506	

表 3.11 修正分類偏誤範例 2

	資料來源	Sensor 1		Sensor 2	
	標準差	5.18		1.81	
明確 分類法		次數	機率	次數	機率
	A	4	0.133	0	0.000
	B	23	0.767	30	1.000
	C	3	0.100	0	0.000
	熵	1.014		0.000	
	權重	---		---	
距離 權重法		次數	機率	次數	機率
	A	3.08	0.103	1.14	0.038
	B	23.85	0.795	27.73	0.924
	C	3.08	0.103	1.14	0.038
	熵	0.937		0.463	
	權重	0.804		0.196	

表 3.12 修正分類偏誤範例 3

	資料來源	Sensor 1		Sensor 2	
	標準差	9.41		15.12	
明確 分類法		次數	機率	次數	機率
	A	10	0.333	11	0.367
	B	9	0.300	5	0.167
	C	11	0.367	14	0.467
	熵	1.580		1.475	
	權重	0.466		0.534	
距離 權重法		次數	機率	次數	機率
	A	6.12	0.204	8.30	0.277
	B	17.76	0.592	12.71	0.424
	C	6.12	0.204	8.99	0.300
	熵	1.384		1.559	
	權重	0.559		0.441	

四、模擬測試

本研究在上一個章節已對資料融合模式的基本架構做介紹，期望以距離權重法來改善最佳權重法的分類機率計算方式，降低上述偏誤的影響，為了了解修正後的結果是否有明顯的改善，因此本研究以電腦模擬的方式，假設可能產生偏誤的情況，分別計算出距離權重法以及明確分類法的結果，加以比較，希望能評估模式改善的效果。本章在 4.1 節說明模擬資料的產生方式，4.2 節對測試情境的設計做一個簡單的描述，4.3 節將模擬測試的結果作分析。

4.1 模擬資料之產生

本研究期望以電腦模擬的方式，來測試距離權重法改善的效果，但由於要以電腦產生模擬的車速資料，必須要先了解車速資料的分佈型態，所以本研究擬用歷史交通資料，以配合度檢定來推估其分配的型式。

在歷史資料方面，採用一計程車車隊的 GPS 資料，分佈的時間及地點如下，車速資料如表 4.1 所示。

日期：94 年 10 月 7 日

時間：19:00~20:00

地點：新竹市東大路二段 武陵路口—水田街口段

首先我們假設母體為常態分配，建立對等假設如下：

H_0 ：母體為常態分配

H_1 ：母體非常態分配

設定型 I 誤差風險值 $\alpha=0.05$ ，並計算相關參數，如表 4.2 所示，並以卡方檢定來看推論是否正確。

表 4.1 GPS 車速資料

時間	車速(kph)	時間	車速(kph)
19:03:12	20.37	20:10:37	18.52
19:07:39	48.15	20:24:13	3.70
19:08:58	5.56	20:24:37	16.67
19:09:43	31.48	20:25:28	9.26
19:10:20	27.78	20:26:58	24.08
19:10:24	35.19	20:32:35	31.48
19:10:40	16.67	20:34:56	7.41
19:11:40	31.48	20:36:35	37.04
19:13:57	9.26	20:36:36	35.19
19:14:23	18.52	20:37:43	14.82
19:15:27	18.52	20:38:35	9.26
19:17:34	25.93	20:43:45	22.22
19:31:11	14.82	20:44:00	31.48
19:34:59	22.22	20:50:22	5.56
19:41:30	27.78	20:50:52	5.56
19:43:28	51.86	20:50:54	11.11
19:48:37	18.52	20:52:23	9.26
19:49:46	29.63	20:53:18	9.26
19:56:13	31.48	20:53:22	18.52
19:58:33	9.26	20:53:48	7.41
19:59:55	27.78	20:53:58	31.48
20:01:03	9.26	20:54:50	14.82
20:02:42	20.37	20:57:57	7.41
20:05:06	29.63	20:59:25	18.52
20:07:15	16.67	20:59:27	7.41
20:09:07	22.22	20:59:57	5.56
20:10:23	1.85		
平均車速	19.53		
標準差	11.48		

表 4.2 卡方檢定頻率分佈計算

車速分類 (kph)	樣本頻率	接受 H_0		相對平方殘餘值
		機率	期望頻率	
0-11	17	0.285	15.11	0.24
11-22	15	0.300	15.90	0.05
22-33	16	0.295	15.64	0.01
33-44	3	0.103	5.46	1.11
44-55	2	0.017	0.90	1.34
總和	53	1.000	53	2.75

由於設定的分類數為 5，常態分配的參數有均數及標準差兩個，因此卡方分配的自由度為 $5-2-1=2$ ，而在 $\alpha=0.05$ 下， $\chi^2(0.95;2)=5.99$ ，判定法則為：

若 $X^2 \leq 9.21$ ，則接受 H_0 ，母體為常態分配。

若 $X^2 > 9.21$ ，則接受 H_1 ，母體非常態分配。

由表 4.2 計算出相對平方殘餘值為 2.75，小於 5.99，因此接受 H_0 ，母體為常態分配。

得知車速資料為常態分配後，本研究以 Visual Basic 程式語言自行撰寫常態分配產生器，來控制母體均數、標準差及樣本數，若產生的車速資料小於 0 或大於 70，則視為極端值刪除，直到產生足夠之樣本數，圖 4.1 為車速資料產生流程圖。

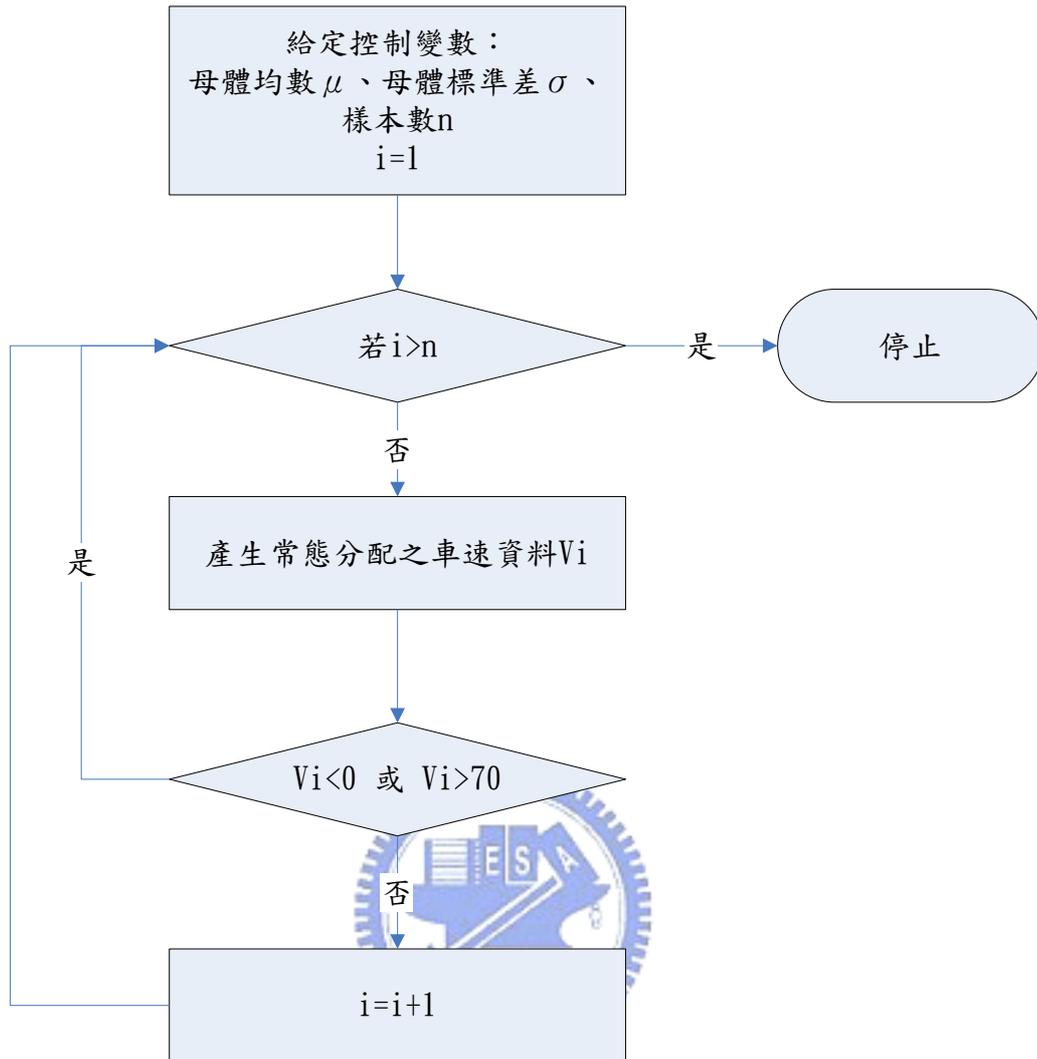


圖 4.1 常態分配產生流程圖

4.2 實驗設計

本研究與過去最佳權重法不同的地方，在於以距離權重法來改善明確集合的分類方式，減少明確分類法所造成的偏誤，因此為了比較距離權重法與明確分類法的差異，在實驗設計上，主要針對可能發生偏誤的情況來做情境的假設。

明確分類法可能發生偏誤的情況有三種：

1. 車速資料大多集中於分類 B，且部分資料於分類邊界附近。
2. 所有資料皆為同一分類。
3. 資料的變異程度過大。

以下則對上述三種情況來做實驗設計：

情境一：為了測試資料分佈集中於分類 B，且部分資料於分類邊界附近的情況，因此設計兩資料的樣本數、母體均數相同，且母體標準差皆為 3，實驗設計如表 4.3。理論上變異程度相同的兩組資料，所得的權重應相差不大，藉由情境一的實驗來觀察以距離權重法是否能改善明確分類法的偏誤。

表 4.3 實驗設計—情境一

資料	樣本數	母體均數	母體標準差
資料一	30	35	3
資料二	30	35	3

情境二：為了觀察所有資料皆為同一分類的情況，故將資料一的標準差設為 2，則所有的資料會分至同一組內，由於此方案無法以明確分類法來計算權重，所以僅觀察當資料二為不同標準差時，距離權重法是否能合理的分配權重。實驗設計如表 4.4。

表 4.4 實驗設計—情境二

資料	樣本數	母體均數	母體標準差
資料一	30	35	2
資料二	30	35	3
資料一	30	35	2
資料二	30	35	4
資料一	30	35	2
資料二	30	35	5
資料一	30	35	2
資料二	30	35	6

情境三：由於當資料變異程度過大時，明確分類法會使得 A、C 兩類機率提高，而熵值下降的情況，為了測試此種偏誤改善的結果，因此控制資料一的母體標準差為 10，改變資料二的母體標準差來觀察權重的變化。實驗設計如表 4.5。

表 4.5 實驗設計—情境三

資料	樣本數	母體均數	母體標準差
資料一	30	35	10
資料二	30	35	13
資料一	30	35	10
資料二	30	35	15
資料一	30	35	10
資料二	30	35	17
資料一	30	35	10
資料二	30	35	19

4.3 測試結果

本節將依照情境一、二、三設計的參數，來產生車速資料，觀察兩種分類方法獲得的權重，並加以分析及比較。

4.3.1 情境一

情境一的計算步驟及結果如表 4.6~4.9 所示。

由表 4.8 可以得知，當資料分佈集中於分類 B 的時候，以明確分類法計算，分類 B 會有相當高的機率函數，而且又因為樣本數不多，每一筆資料的分類對於機率函數的敏感度很高，因此只要在分類次數上相差一筆，就會容易造成兩組資料的不確定性差異很大，而求得的權重也相去甚遠。如情境一的結果所示，兩組資料的母體標準差相同，但以明確分類法計算得的權重，其中一組達七成左右，而第三組的測試範例竟高達九成。以距離權重法計算，權重較不會受到分類邊界的影響，兩組資料的權重皆在 0.4~0.6 之間。

表 4.6 模擬車速資料—情境一

資料	樣本數	樣本均數	樣本標準差
資料一	30	35.2	3.07
資料二	30	35.5	2.88
資料一	30	35.2	2.91
資料二	30	35.5	2.78
資料一	30	34.5	2.79
資料二	30	36.2	2.78
資料一	30	35.7	2.90
資料二	30	34.8	2.79

表 4.7 分類次數—情境一

資料	明確分類法			距離權重法		
	C	B	A	C	B	A
資料一	3	26	1	1.79	26.42	1.79
資料二	2	28	0	1.76	26.48	1.76
資料一	1	28	1	1.78	26.44	1.78
資料二	0	29	1	1.70	26.61	1.70
資料一	2	26	2	1.69	26.62	1.69
資料二	0	29	1	1.62	26.76	1.62
資料一	1	28	1	1.84	26.32	1.84
資料二	1	29	0	1.69	26.62	1.69

表 4.8 分類機率—情境一

資料	明確分類法			距離權重法		
	C	B	A	C	B	A
資料一	0.100	0.867	0.033	0.060	0.881	0.060
資料二	0.067	0.933	0.000	0.059	0.883	0.059
資料一	0.033	0.933	0.033	0.059	0.881	0.059
資料二	0.000	0.967	0.033	0.057	0.887	0.057
資料一	0.067	0.867	0.067	0.056	0.887	0.056
資料二	0.000	0.967	0.033	0.054	0.892	0.054
資料一	0.033	0.933	0.033	0.061	0.877	0.061
資料二	0.033	0.967	0.000	0.056	0.887	0.056

表 4.9 熵與權重—情境一

資料	明確分類法		距離權重法	
	熵	權重	熵	權重
資料一	0.675	0.225	0.647	0.494
資料二	0.363	0.775	0.640	0.506
資料一	0.420	0.217	0.644	0.483
資料二	0.221	0.783	0.622	0.517
資料一	0.700	0.091	0.621	0.485
資料二	0.221	0.909	0.602	0.515
資料一	0.420	0.217	0.659	0.469
資料二	0.221	0.783	0.620	0.531

4.3.2 情境二

情境二的計算步驟及結果如表 4.10 ~ 4.13 所示。

由此結果可知，當車速資料皆於同一分類的時候，距離權重法仍可計算兩資料的權重，且資料二的權重會隨著其標準差的提高而降低。另外，當資料二的標準差為 4.23 時，資料一的權重已達 0.81，其原因在於以 3.11 式計算的權重，除了與各資料的熵值有關以外，也和資料的熵值比有關，因此當兩資料的不確定性較低時，熵值的差異對於權重的影響就會提高。由於情境二的標準差皆較情境一小，如上述所言，同以距離權重法的結果來看，情境二的標準差對於權重的敏感度較情境一來的高。

表 4.10 模擬車速資料—情境二

資料	樣本數	樣本均數	樣本標準差
資料一	30	35.1	1.99
資料二	30	34.9	3.20
資料一	30	35.6	1.81
資料二	30	33.8	4.23
資料一	30	34.9	1.81
資料二	30	32.8	5.18
資料一	30	35.0	1.80
資料二	30	36.3	6.40

表 4.11 分類次數—情境二

資料	距離權重法		
	C	B	A
資料一	1.22	27.56	1.22
資料二	1.84	26.32	1.84
資料一	0.97	28.05	0.97
資料二	2.65	24.70	2.65
資料一	1.14	27.73	1.14
資料二	3.08	23.85	3.08
資料一	1.14	27.73	1.14
資料二	3.84	22.32	3.84

表 4.12 分類機率—情境二

資料	距離權重法		
	C	B	A
資料一	0.041	0.919	0.041
資料二	0.061	0.877	0.061
資料一	0.032	0.935	0.032
資料二	0.088	0.823	0.088
資料一	0.038	0.924	0.038
資料二	0.103	0.795	0.103
資料一	0.038	0.924	0.038
資料二	0.128	0.744	0.128

表 4.13 熵與權重—情境二

資料	距離權重法	
	熵	權重
資料一	0.488	0.646
資料二	0.659	0.354
資料一	0.411	0.810
資料二	0.849	0.190
資料一	0.463	0.804
資料二	0.937	0.196
資料一	0.463	0.844
資料二	1.076	0.156

4.3.3 情境三

情境三的計算步驟及結果如表 4.14~4.17 所示。

由此結果可知，以明確分類法來統計次數，當標準差約為 10 的時候，A、C 類的次數就會大於 B 類，因此若再將標準差加大，則熵反而會降低，權重會隨之提高；以距離權重法來計算，標準差要提高到約為 17 時，A、C 類的次數才會大於 B 類，因此可降低資料變異程度過大所造成偏誤的影響。但是，由於此情境的資料變異程度皆較大，因此資料對於權重的敏感度會過小，以致於雖然權重會隨著標準差的提高而降低，然而降低的幅度並不明顯。

另外，會造成此種偏誤的原因在於 B 類的區間較 A、C 兩類小，所以當資料變異程度大時，A、C 兩類的機率函數會大於 B 類，因此若將分類的區間修正為等距，也可以降低此種偏誤的影響。

表 4.14 模擬車速資料—情境三

資料	樣本數	樣本均數	樣本標準差
資料一	30	39.5	10.12
資料二	30	38.0	13.52
資料一	30	37.4	9.41
資料二	30	35.5	15.12
資料一	30	39.6	10.59
資料二	30	41.0	17.52
資料一	30	35.3	9.15
資料二	30	36.0	20.46

表 4.15 分類次數—情境三

資料	明確分類法			距離權重法		
	C	B	A	C	B	A
資料一	10	11	9	6.18	17.64	6.18
資料二	10	9	11	7.31	14.88	7.81
資料一	11	9	10	6.12	17.76	6.12
資料二	14	5	11	8.99	12.71	8.30
資料一	8	13	9	5.91	17.91	6.18
資料二	12	4	14	9.24	9.75	11.01
資料一	9	12	9	5.80	18.40	5.80
資料二	14	3	13	11.75	7.81	10.44

表 4.16 分類機率—情境三

資料	明確分類法			距離權重法		
	C	B	A	C	B	A
資料一	0.333	0.367	0.300	0.206	0.588	0.206
資料二	0.333	0.300	0.367	0.244	0.496	0.260
資料一	0.367	0.300	0.333	0.204	0.592	0.204
資料二	0.467	0.167	0.367	0.300	0.424	0.277
資料一	0.267	0.433	0.300	0.197	0.597	0.206
資料二	0.400	0.133	0.467	0.308	0.325	0.367
資料一	0.300	0.400	0.300	0.193	0.613	0.193
資料二	0.467	0.100	0.433	0.392	0.260	0.348

表 4.17 熵與權重—情境三

資料	明確分類法		距離權重法	
	熵	權重	熵	權重
資料一	1.580	0.500	1.390	0.539
資料二	1.580	0.500	1.503	0.461
資料一	1.580	0.466	1.384	0.559
資料二	1.475	0.534	1.559	0.441
資料一	1.552	0.459	1.376	0.569
資料二	1.429	0.541	1.581	0.431
資料一	1.571	0.431	1.349	0.574
資料二	1.368	0.569	1.565	0.426

4.4 模擬測試小結

由上述各方案的結果，我們可以得到幾點結論：

1. 當資料分佈集中於分類 B 的時候，以明確分類法計算，會因為資料對機率函數的敏感度過大，而造成過度信賴某組資料的情況。以距離權重法計算，權重較不會受到分類邊界的影響。
2. 當車速資料皆於同一分類時，用明確分類法無法計算權重；以距離權重法則可計算，且權重會依標準差增大而變小，但由於一組資料的不確定性過小，而導致此種情況資料變異程度對權重的敏感度提高。
3. 當資料的變異程度較大時，以明確分類法來計算，當標準差約為 10 的時候，熵值就會開始降低，而造成偏誤；以距離權重法來計算，標準差要提高到約為 17 時，熵值才會開始降低，因此可降低資料變異程度過大所造成偏誤的影響。此外，也可將各分類的區間修正為等距，來降低資料變異程度大的影響。
4. 最佳權重法計算的權重會與各資料的熵值比有關，也就是說當所有資料的熵值皆較大時，熵值的差異對於權重的影響會降低；而當某一組資料的熵值很小時，則熵值對權重的敏感度會提高。

五、結論與建議

本研究期望能發展資料融合模式，用以整合交通資料，提供較可靠的即時路況資訊。最佳權重法能將異常的觀測值表現在熵值上，且不會受到資料平均值的影響，唯其受到資料分類的限制，因此提出距離權重法來改善資料分類造成的偏誤，根據測試的結果，可以歸納出以下的結論與建議。

5.1 結論

1. 最佳權重法能反應資料的變異程度，但無法得知樣本均數與實際值的差異，也就是說此模式僅考慮資料的可靠度，而不考慮資料的準確度。
2. 最佳權重法計算的權重會與各資料的熵值比有關，也就是說當所有資料的熵值皆較大時，熵值的差異對於權重的影響會降低；而當某一組資料的熵值很小時，則熵值對權重的敏感度會提高。
3. 距離權重法的分類方式不致於造成過度信賴某組資料的情況，且當資料相當集中的時候，也可計算出各資料的權重。
4. 由於各分類區間不相同，同資料變異程度大時，熵值反而會下降，若將各分類區間修正為等距，可改善此種偏誤的影響。
5. 由於熵值對於權重的敏感度會隨著熵值的提高而降低，因此若有一組資料的熵值過大時，反而會因為敏感度過小，使得不論另一組資料的熵值為何，其計算出的權重皆會相近，無法區別其差異性，因此本研究提出之方法，適合用於各組資料的不確定性不致於太高的情況下。

5.2 建議

1. 本研究的資料分類方式是以各組的組中點當種子，來計算每一筆資料分類的權重，因此計算結果會依組種子的不同而有所改變，後續可以針對不同組種子的取決方式做探討。

2. 本研究僅以資料的可靠度來做為資料融合的依據，建議未來可再發展以資料的準確度為基礎的資料融合模式。



六、參考文獻

- [1] Andrej Rakar, Dani Juricic, Peter Ballé, “Transferable Belief Model in Fault Diagnosis”, *Engineering Applications of Artificial Intelligence*, 12, pp. 555-567, 1999.
- [2] Benoît Duc, Elizabeth Saers Bigün, Josef Bigün, Gilbert Maître, Stefan Fischer, “Fusion of Audio and Video Information for Multi Modal Person Authentication”, *Pattern Recognition Letters*, 18, pp. 835-843, 1997.
- [3] Daniel J. Dailey, Patricia Harn, Po-Jung Lin, “The Final Research Report of ITS Data Fusion”, Washington State Transportation Center and Washington State Department of Transportation, 1996.
- [4] Hall D, L., Linn R. j., “A Taxonomy of Algorithm for Multi-sensor Data Fusion”, *Technical Proceedings of the Joint service Data Fusion Symposium*, Vol. I, pp. 594-610, 1990.
- [5] Huadong Wu, Mel Siegel, Rainer Stiefelhagen, Jie Yang, “Sensor Fusion Using Dempster-Shafer Theory”, *IEEE Instrumentation and Measurement Technology Conference*, 2002.
- [6] Huadong Wu, Mel Siegel, Sevim Ablay, “Sensor Fusion Using Dempster-Shafer Theory II: Static Weighting and Kalman Filter-like Dynamic Weighting” *IEEE Instrumentation and Measurement Technology Conference*, 2003.
- [7] Keechoo Choi, YounShik Chung, “A Data Fusion Algorithm for Estimating Link Travel Time”, *Intelligent Transportation Systems*, 7, pp. 235-260, 2002.
- [8] Klein Lawrence A., “Sensor Technologies and Data Requirements for ITS” Artech House, 2001.
- [9] Linn R. J., D. L. Hall, “A Survey of Multi-sensor Data Fusion Systems” *Proceedings of the SPIE-The International Society for Optical Engineering*, Vol 1470, pp. 13-29, 1991.
- [10] Otman A. Basir, Helen C. Shen, “Sensory Data Integration: A Team Consensus Approach”, *IEEE Robotics and Automation Conference*, pp. 1683-1688, 1992.
- [11] Ruey Long Cheu, Der-Horng Lee, Chi Xie, “An Arterial Speed Estimation Model Fusing Data from Stationary and Mobile Sensors”, *IEEE Intelligent Transportation Systems Conference*, pp. 573-578, 2001.
- [12] Sarma V. S., S. Raju, “Multisensor Data Fusion and Decision Support for Airborne Target Identification”, *IEEE Transactions on Systems, Man and Cybernetics*, Sept.-Oct., 1991.
- [13] 朱良浩, 「地理資訊系統在區域合併問題上之應用」, 國立交通大學碩士論文, 民國八十三年。
- [14] 吳欣潔, 「熵應用於交通資料融合之研究」, 國立交通大學碩士論文, 民國

九十三年。

- [15] 曾治維，「Dempster-Shafer 理論於交通資料整合技術之應用」，國立交通大學碩士論文，民國九十三年。

