

# 行政院國家科學委員會專題研究計畫 期中進度報告

用路人異質資訊融合與應用技術平台之開發--子計畫一：  
異質交通運輸資訊融合技術之研究(2/3)  
期中進度報告(精簡版)

計畫類別：整合型  
計畫編號：NSC 94-2218-E-009-012-  
執行期間：94年08月01日至95年07月31日  
執行單位：國立交通大學運輸科技與管理學系(所)

計畫主持人：王晉元

處理方式：本計畫可公開查詢

中華民國 95 年 12 月 25 日

# 行政院國家科學委員會補助專題研究計畫期中進度報告

用路人異質資訊融合與應用技術平台之開發-子計畫一：異質交通運輸資訊融合技術之研究(II)

計畫類別： 個別型計畫  整合型計畫

計畫編號：NSC 94－ 2218－E－009－ 012－

執行期間： 94 年 8 月 1 日至 95 年 7 月 31 日

計畫主持人：王晉元

共同主持人：

計畫參與人員： 吳瑞豐、陳威豪

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：國立交通大學 運輸科技與管理學系

中 華 民 國 95 年 5 月 1 日

# 最佳權重法應用於交通資料融合

## 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，測試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。

**關鍵詞：**旅行者資訊、資料融合、先進旅行者資訊系統、熵

## 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

## 一、緒論

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。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。

然而即時路況資訊可由許多來源獲得，如固定式車輛偵測器、GPS 探針車等，各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，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。

資料融合包括資料的收集、過濾、分析以及結合，本研究的重點在於多種異質資料的結合，整合數個來源的交通資料。在過去學術研究中，資料融合在各個領域的發展相當廣泛。Otman and Helen [1992]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，以熵來表示系統的不確定性，使得融合後的結果具有最小的熵，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。

Andrej et al. [1999]提出以 TBM (Transferable Belief Model)來找出生產線上的錯誤，並以直流馬達的錯誤檢測作為測試，TBM 的概念與傳統機率理論相似，但是 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。研究的結果認為 TBM 容易與模糊邏輯結合應用，且正確率優於布林推論(Boolean Reasoning)的結果。

Ruey et al. [2001]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。在測試範例中，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，但卻無法修正車速因車輛停等紅綠燈所造成的誤差。

Huadong et al. [2002]提出以權重 Dempster-Shafer 理論，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。在測試範例中比較了線性加總、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。曾治維[2004]將 Dempster-Shafer 理論應用於交通資料整合上，此研究結果發現若在誤差較大的情況下，融合後的結果則較不盡理想。

Keechoo and YounShik [2002]利用模糊回歸來融合 GPS 探針車與固定式車輛偵測器的資料，Keechoo and YounShik 用 MAPE 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。

總體來說，TBM 雖然能夠降低傳統機率理論的偏誤，但交通資料出現的情況是可預測的，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。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，然而交通資料的收集實屬不易，無法用大量的資料來校估參數，容易影響準確度。團隊共識法能將異常的觀測值表現在熵值上，唯其受到資料分佈位置及分類區間的限制。因此本研究期望修正團隊共識法之分類方式，發展出一套較合適交通資料的融合模式，並於第二部分介紹

模式的架構，在第三部分則以模擬的車速資料來做測試，最後再依據測試的結果作一個綜合的評論。

## 二、模式構建

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(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  $V$ 。

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

由(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。最佳權重法為修正團隊共識法之交通資料分類方式，其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。本研究以 Shannon 熵來代表各偵測器的不確定性，Shannon 熵為信息理論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。

### 2.1 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

$$T = \sum_{i \in S} H_i \quad (2)$$

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

$$\begin{aligned} \text{Minimize} \quad & \sum_{i \in S} W_i^2 H_i^2 \\ \text{Subject to} \quad & \sum_{i=1}^N W_i = 1 \\ & W_i > 0 \end{aligned} \quad (3)$$

(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，而其限制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。

由(3)式可知，權重  $W_i$  為不確定性  $H_i$  的函數，因此只要能定義資料的不確定性  $H_i$ ，即可求得權重  $W_i$ 。

## 2.2 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：

對固定  $n$  來說， $H$  是  $p_1, p_2, \dots, p_n$  的連續函數。

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。

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：

$$H(p_1, p_2, \dots, p_n) = -\sum_{i=1}^n p_i \log_2 p_i \quad (5)$$

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。

### 2.3 最佳權重法應用於交通資料融合

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(5)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，每個等級之服務水準各有不同的標準。為了簡化模式，則將資料分類修正為三類，如表 1 所示。

表 1 車速分類

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 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 2。

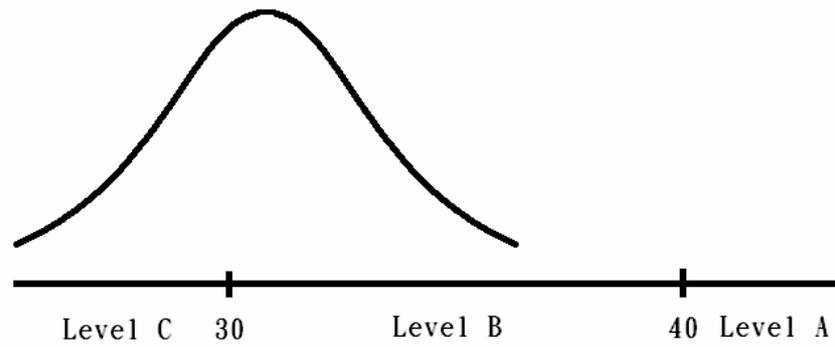


圖 1 資料分佈於分類邊界示意圖

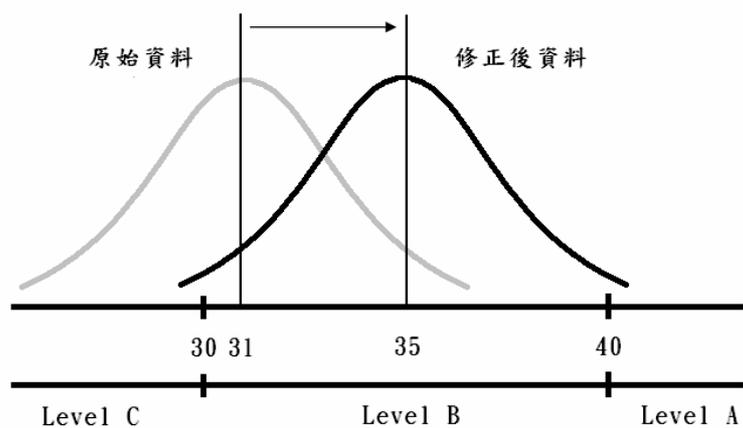


圖 2 資料平移示意圖

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。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，如表 2 所示，並依據各分類的機率函數，計算此組資料的熵為 1.524。

表 2 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

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。表 3 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 最佳權重法範例

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	

## 2.4 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，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、表 4 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。

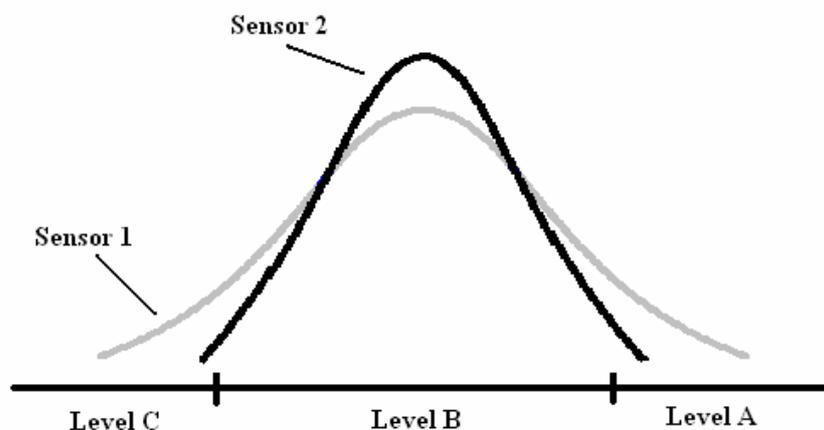


圖 3 分類偏誤示意圖 1

表 4 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	<b>0.675</b>		<b>0.363</b>	
權重	<b>0.225</b>		<b>0.775</b>	

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 4、表 5 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。

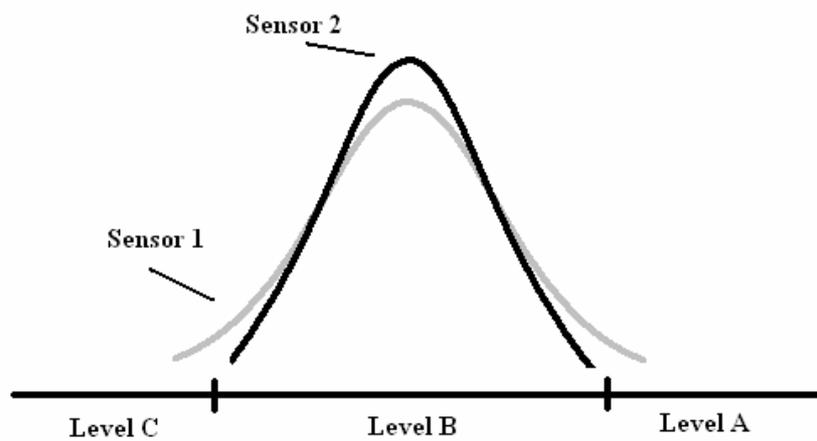


圖 4 分類偏誤示意圖 2

表 5 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	<b>1.014</b>		<b>0.000</b>	
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 5、表 6 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。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。

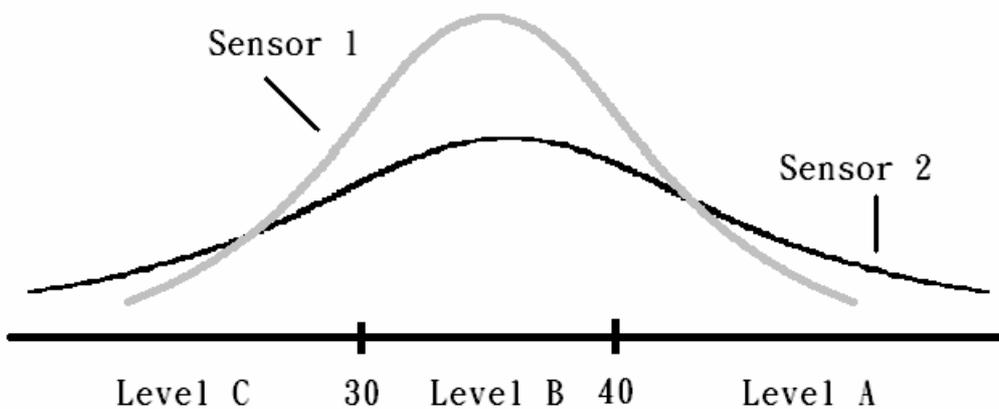


圖 5 分類偏誤示意圖 3

表 6 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	<b>1.580</b>		<b>1.475</b>	
權重	<b>0.466</b>		<b>0.534</b>	

由於依上述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，如下列所述。

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 7，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

$$N_B = \frac{1}{2.2(\frac{1}{17.8} + \frac{1}{2.2})} = 0.89, \quad N_C = \frac{1}{17.8(\frac{1}{17.8} + \frac{1}{2.2})} = 0.11, \quad \text{且分配之次數總和為 1。}$$

表 7 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 4、5、6 三例經修正後，如表 8、9、10 所示。由表 8 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 9 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 10 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 8 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		次數	機率	次數	機率

分類法	A	1	0.033	0	0.000
	B	26	0.867	28	0.933
	C	3	0.100	2	0.067
	熵	<b>0.675</b>		<b>0.363</b>	
	權重	<b>0.225</b>		<b>0.775</b>	
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	<b>0.647</b>		<b>0.640</b>	
	權重	<b>0.494</b>		<b>0.506</b>	

表 9 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	<b>1.014</b>		<b>0.000</b>	
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	<b>0.937</b>		<b>0.463</b>	
權重	<b>0.804</b>		<b>0.196</b>		

表 10 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		次數	機率	次數	機率

分類法	A	10	0.333	11	0.367
	B	9	0.300	5	0.167
	C	11	0.367	14	0.467
	熵	<b>1.580</b>		<b>1.475</b>	
	權重	<b>0.466</b>		<b>0.534</b>	
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	<b>1.384</b>		<b>1.559</b>	
權重	<b>0.559</b>		<b>0.441</b>		

### 三、模擬測試

本研究的重點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，本研究以電腦模擬的方式，假設車速資料為常態分配，控制母體的均數及標準差，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。

第二種情況已於第二部分的範例證實，當所有資料皆為同一分類時，以距離權重法還是可以計算各偵測器之權重，而第三種情況則以等區間的方式修正也可改善，因此以下針對第一種情況來做實驗設計。

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。理論上變異程度相同的兩組資料，所得的權重應相差不大，藉由模擬測試來觀察以距離權重法是否能改善明確分類法的偏誤。

測試結果如表 11-14 所示。由表 13 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。結果顯示，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 11 模擬車速資料

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 12 分類次數

資料	固定分類法			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 13 分類機率

資料	固定分類法			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 14 熵與權重

資料	固定分類法		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

由上面的測試結果，我們可以發現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。

## 四、結論

本研究利用最佳權重法來發展資料融合模式，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，根據測試的結果顯示，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。而以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。

此外，由於各分類區間不相同，當資料變異程度大時，熵值反而會下降，建議將各分類區間修正為等距，可改善此種偏誤的影響。本研究僅以資料的可靠度來做為資料融合的依据，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。

## 參考文獻

1. 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。
2. Andrej Rakar, Dani Juricic, Peter Ballê, “Transferable Belief Model in Fault Diagnosis”, *Engineering Applications of Artificial Intelligence*, 12, pp. 555-567, 1999.
3. 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.
4. 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.
5. 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.
6. Huadong Wu, Mel Siegel, Rainer Stiefelhagen, Jie Yang, “Sensor Fusion Using Dempster-Shafer Theory”, *IEEE Instrumentation and Measurement Technology Conference*, 2002.
7. 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.
8. Keechoo Choi, YounShik Chung, “A Data Fusion Algorithm for Estimating Link Travel Time”, *Intelligent Transportation Systems*, 7, pp. 235-260, 2002.
9. Klein Lawrence A., “Sensor Technologies and Data Requirements for ITS” Artech House, 2001.
10. 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.
11. Otman A. Basir, Helen C. Shen, “Sensory Data Integration: A Team Consensus Approach”, *IEEE Robotics and Automation Conference*, pp. 1683-1688, 1992.
12. 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.
13. 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.

## 計畫成果自評

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，測試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。

本研究的成果已經撰寫成論文，目前正在投稿審查中。

本研究的成果，除了具有學術性外，對於國內旅行者資訊的提供亦具有顯著的貢獻。

本研究的成果大致與原先預定的工作項目相符。