

行政院國家科學委員會專題研究計畫 成果報告

應用空間資訊技術與混合數值算則於分佈型水文系統之模擬 研究成果報告(精簡版)

計畫類別：個別型
計畫編號：NSC 97-2221-E-040-007-
執行期間：97年08月01日至98年10月31日
執行單位：中山醫學大學應用資訊科學學系(所)

計畫主持人：曾明性

計畫參與人員：碩士班研究生-兼任助理人員：湯秉宏

處理方式：本計畫可公開查詢

中華民國 99 年 01 月 29 日

一、緣由與目的

水文學為地球科學中相當重要的一門科學，而水文模式因可對水文歷程進行動態模擬，常為進行水文系統相關研究的重要工具。然而，每個模式因描述水文現象的隨機性、空間分佈與時間變異等特性有所不同而有相當程度的差異。根據 Chow et al. (1988) 的定義，水文模式可概分為兩大類：定率型模式(Deterministic Models)與序率型模式(Stochastic Models)，定率型模式著重在現象的分析，而序率型模式則偏重在時間序列的隨機處理。其中，定率型模式又分不考慮空間變化的集塊型模式(Lumped Models)和考慮空間變化的分佈型模式(Distributed Models)。分佈型降雨逕流模式主要探討集水區水文特性與地文特性對逕流歷線之影響，可適用於土地利用發生變化或無水文紀錄地區，進行集水區水文反應之推估(Beven and Ó Connell, 1982)。

近年來，隨著資訊科技日新月異，電腦計算能力大幅提升，各種航遙測(Remote Sensing, RS)影像技術的發達，與可整合處理大量空間資料之地理資訊系統(Geographic Information System, GIS)的成熟，促成數值地形模型(Digital Terrain Model, DTM)、土地利用、植生覆蓋、土壤含水量等…環境資料庫相繼建立，而其資料精度亦不斷增加，這些資料將有助於水文模式參數之獲取，亦造就高精度之分佈型水文模式的開發成為研究趨勢。

本計畫原先規劃研究目的擬以三年期程，運用申請人研發的高解析顯隱混合有限體積算則整合空間資訊學(地理資訊系統、數值地形模型、遙測影像資料探勘)的處理技術，構建一個新的分佈物理型水文模式，期能針對單場洪水事件之集水區降雨逕流的時空分佈進行更精確的預測模擬。但由於國科會僅核給一年期計畫，故本成果報告以能獨立運作的多光譜影像判識模組的開發工作成果為本報告內容。

二、文獻回顧

90 年代以來，隨著電腦運算技術之快速進步，地理資訊系統(Geographic Information System, GIS)、數值地形模型/數值高程模型(Digital Terrain Model/Digital Elevation Model, DTM/DEM)、與遙感探測(Remote Sensing, RS)等技術迅速發展，可輔助求取分佈物理型降雨逕流模式中所需

之各項地文及水文因子，為研究與建立分佈型水文模式提供了有利技術及資料上的支撐，亦造就分佈型水文模式之研究遂成為當前集水區降雨逕流關係推演之重要課題。本期計畫主要工作內容為多光譜影像判識模組的開發，以下茲就遙測影像分類之國內外研究情形與重要前人文獻分項說明如下：

遙測影像具有大地區的空間涵蓋特性，可以提供詳實且長時間的數位化記錄，並具有極佳的觀察能力，是迅速獲得全面性及即時的土地利用監測工具與最新國土資訊的調查工具。近年來，由於遙測相關技術蓬勃發展，利用遙測影像進行地表覆蓋或土地利用的自動判釋技術，在分佈型水文模式的發展上佔有基本且重要的一環。遙感探測的影像可分單一光譜(Panchromatic)影像、多光譜(Multispectral)影像(僅 3—50 譜帶數，光譜解析度 100 nm)、及高光譜(Hyperspectral)影像(譜帶數高達 100—250，光譜解析度達 3nm)。多(高)光譜遙測影像即是以測量各光譜帶電磁波的反射值來描述地表特徵，由於不同物種所反射之電磁波波長並不相同，因此可藉由電磁波之反射波長推算集水區之地表覆蓋或土地利用情形。透過遙測影像自動分類可監測集水區土地使用或地表覆蓋的改變情況，水文學家能夠更深入瞭解環境變遷的現況，藉以提高集水區降雨逕流模擬技術的精進。

一般而言，光譜影像分類方法可分為監督式學習法(Supervised Learning)與非監督式學習法(Unsupervised Learning)，其中監督式學習法較準確而受到廣用。傳統上常用的統計技術如較小距離或最大概似分類器(Maximum Likelihood Classifier, MLC) (李光敦,2006；王毓麒, 2006)，將各分類物種假設為高斯常態分佈，但常常各類別實際上並不符合常態分佈，而造成分類錯誤(Tseng et al., 2007)。近年來根源於人工智慧的影像判識技術遂成為主要研究方向(Openshaw and Openshaw, 1997; Tseng et al., 2008)。王乃宣與曾明性(2006)以兩組具有差異性之多光譜遙測影像為測試樣本，進行包含 kNN、k-means+kNN、k-medoids+kNN、GA+kNN、SVM 等五種分類演算法的效能評估，經由自行開發影像分類器程式，希望能藉由遙測影像達成地表覆蓋或土地使用之智慧型自動分類所需，克服現地調查或人工判識在成本與時效上的缺失。由於目前影像分類研究最常使用的分類技術如簡單貝氏分類、類神

經網路、最近芳鄰法、支持向量機等都是以黑盒方式進行影像分類，使用者並無法瞭解其中隱含的知識邏輯。Tseng et al. (2008)成功開發基因演算規則式分類器，除可自動判識各種地表覆蓋類型外，還可直接進行分類知識規則的擷取，可增進人類對影像分類的知識學習外，更利於提升後續大量影像資料的判識效能。

三、研究方法

1. 屬性篩選

影像分類是資料探勘技術的一種應用，首要任務是選擇輸入變數，也就是資料維度精簡或屬性篩選(Feature Selection)；當遇到資料集有相當多的屬性，但並非每個屬性都會影響到分類結果，亦即某些屬性可能對目標值的影響相當小時，屬性篩選可以去除這些不重要的屬性以提高資料探勘的正確率(曾明性等, 2007; 張彙音與曾明性等, 2006)。資料精簡技術的方法有逐步回歸法(Stepwise Regression)、因素分析(Factor Analysis)、主成分分析(Principal Components Analysis)、相關性檢定(Correlation Test)、資料測量法(Data Measurement)、特徵權重法(Feature Weighting)等。申請人以往使用資訊量測量法、數值型屬性測量法、ESX 數值屬性測量法、權重型基因演算法、篩選型基因演算法等方法進行屬性探勘(Tseng et al., 2008; 曾明性等, 2007; 張彙音與曾明性等, 2006; 陳盛哲與曾明性, 2006a; 2006b)來進行多光譜影像的屬性篩選。

2. 分類技術

本研究擬使用資料探勘監督式學習法之分類原理，經由多光譜影像的光譜資料庫進行地表覆蓋或土地使用類型的自動判識。以下簡要介紹本研究預計評估的分類技術：最近芳鄰法(k-Nearest Neighbors)、決策樹(Decision Tree)、單純貝氏分類(Naïve Bayesian Classification)、類神經網路(Artificial Neural Networks)、支援向量機(Support Vector Machines)、基因演算法(Genetic Algorithms)等方法(Tseng et al., 2007; 曾明性等, 2007; 王乃宣與曾明性, 2006; 張彙音與曾明性等, 2006; 陳盛哲與曾明性, 2006a; 2006b)，進行多光譜影像自動分類模組的開發測試。

2.1 最近芳鄰法(K-nearest Neighbor, kNN)

kNN 是資料探勘中常用到的一種簡單分類工具，以目前既有的所有範例為依據，並根據事先設定好的最近鄰居數(k 值)，透過投票的方式，來決定新資料所屬類別演算法，也就是以前 k 個最靠近的鄰居來投票決定自己的類別。

2.2 決策樹(Decision Tree, DT)

決策樹演算法是一種以遞迴方式把資料集合切分成越來越具同質性次集合的分類演算法，其模式結構以樹狀展開，子節點為切割的資料集，終端節點又稱葉節點，並以能取得最高增益比值的節點來做為樹狀結構成長分支的依據，所產生的樹狀結構相較於其他演算法，比較容易讓人理解，這一點其實是相當重要的，因為在進行研究時，有時候後不只是希望有準確的模型就好，對決策者來說，更重要的是能否從規則的內容中獲得啟發，此時，規則是否能夠以人類所能理解的形式呈現就顯得相當重要了。決策樹分岔的準則是根據增益比值 (Gain Ratio)來計算，比較各個屬性值的增益比值，挑選具有最高增益比值的因子來當做分岔依據。

2.3 單純貝氏分類(Naïve Bayes Classifier, NBC)

單純貝氏分類器是一種簡單且實用的分類方法。單純貝氏分類器主要是根據貝氏定理 (Bayesian Theorem)，交換事前 (prior) 及事後 (posteriori) 機率，配合決定分類特性的各屬性彼此間是互相獨立的 (conditional independence) 的假設，來預測分類的結果。只要單純貝氏分類器所涉及的資料屬性，彼此間互相獨立的條件被滿足時，單純貝氏分類器的所得到的最大可能分類結果，與貝氏定理的最大可能假說具有相同的功效。

2.4 類神經網路(Artificial Neural Networks, BPN)

類神經網路是由簡單、高度連結性的人工神經元所構成的電腦處理系統。每個神經元都具有簡單的計算能力。典型的類神經網路分成三層；輸入層、隱藏層、輸出層，各層之間以神經鍵(Synapse)聯結，其輸出值與輸入值的關係式，可以使用輸入值的加權乘積和之函數作為關係式。其中 模仿神經元模型的輸出訊號。f(函數) 模仿神經元模型的轉換函數，用以將輸入單元的輸入值之加權乘積和轉換成處理單元輸出值的數學公式。 W_{ij} 模仿神經

元模型的神經節強度，又名聯結加權值。 W_{ij} 用以表示第 i 個處理單元對第 j 個處理單元之影響強度。 X_i 模仿神經元模型的輸入訊號。 e_j 模仿神經元模型的閾值。類神經網路的運作包含兩個程序學習 (Earning) 與回想 (Recalling)。學習是為了神經元之間的聯結權重調整至能夠配合出現於輸入單元的資料；回想是類神經網接收到訊號後，根據神經的佈署與突觸的權重而產生適當的反應。

2.5 支持向量機(Support Vector Machines, SVM)

支持向量機簡稱 SVM，早在 1970 年開始，由學者 Vladimir Vapnik 研究的一種分類方法。90 年代後才開始漸漸被廣為探討，經過許多學者研究後，在不同領域實際運用分類上皆有不錯的效果，現在已成為機器學習領域很熱門的一種方法。SVM 為建立在機器學習理論的結構風險最小化原則之上，其主要思想是針對二元分類問題，在高維度空間中尋找一個超平面作為二類的分割，以保證最小的分類錯誤率，且 SVM 一個重要的優點就是能處理線性不可分的情況，但是它的一個明顯缺點是訓練預測模組的時間可能會很長。SVM 利用目前現有的資料作訓練，再利用分析這些訓練資料選出幾個支援向量(Support Vector)來代表整體的資料，並將少部份極端值事先剔除，然後將所挑選的支援向量包裝成模型。簡而言之，當用來建立分類模型的資料比數足夠多的時候，分類模型的代表性就更具有意義，可以用來預測下一筆未分類的資料應該屬於哪一類別較適當。

3. 分群技術

本研究亦使用資料探勘非監督式學習法之分群原理，經由多光譜影像的光譜資料庫進行地表覆蓋或土地使用類型的自動判識。以下簡要介紹本研究評估的分群技術：K-Means 法、Fuzzy C-Means 法、ISODATA 法，進行多光譜影像自動分類模組的開發測試。

3.1 K-Means 法

k-means 在資料探勘的領域當中，是一種非監督式學習演算法，常用來解決分群問題。這是一個能使資料物以類聚的重要方法，主旨在於促使有相似特性的資料集結成群，而這些有相似特性的資料成群集合即稱為叢聚

(cluster)，其中又以 K-means 演算法最被普遍應用，此法應用的基礎理論是使用資料點到叢聚中心點的最小平方距離來求分類。K-means 演算法比起其他分群演算法，其它的優點在於簡易的計算，可以快速算出分群結果。

3.2 Fuzzy C-Means 法

模糊 c 平均分群演算法(Fuzzy c-mean clustering method, FCM) 是 Bezdek 於 1973 年提出，此演算法是根據 c-means 分群法所衍生。透過模糊邏輯(Fuzzy Logic) 的概念，進一步提升分群效果。FCM 與 K-means 分群法最大差異在於加入了「模糊」的概念，使得個體不再「絕對」地屬於任何群集，而是以一個介於 0~1 之間的數字來表示該個體隸屬於某個群集的程度。每個個體隸屬於每個群集的程度多寡稱為隸屬函數(Membership Function, MF)，代表個體隸屬於群中心的程度。MF 值越高即代表該個體隸屬於該群集的程度越高。且每一個個體對所有群中心的隸屬函數值總和為 1。

3.3 ISODATA 法

ISODATA 集群分類法是依據像元至各集群中心平均值的光譜距離代表像元和集群特徵的相似程度，當距離最小時表示像元與集群相似度最大。開始計算時可任意給定數個集群中心，然後計算每一像元至各集群中心的光譜距離，若某像元至某集群中心的距離最短，則歸於某集群，而後根據所分出的集群重新計算每個集群中心的平均值，如此不停迭代計算，直到滿足所給定的限制條件為止。其所需設定之計算參數為：(一) 集群數，分析時將依據集群數進行各集群平均向量與標準差之計算，且其必須大於等於土地覆蓋類別數。(二) 最大迭代次數，目的為避免無止盡的重複計算。(三) 收斂門檻值，為兩次迭代計算間，影像所有像元之歸屬集群並未改變之最大比率。若超過，則停止計算。

4. 資料集介紹

本文藉由五個多高譜的衛星影像資料集，探討決策數(DT)、單純貝氏分類(NBC)、類神經網路(BPN)、支持向量機(SVM)、最近芳鄰法(KNN)、模糊最近芳鄰法(Fuzzy-KNN)等六種分類演算法；及 K-Means 法、Fuzzy C-Means 法、ISODATA 法等三種分群演算法之比較。五個資料集的基本資

料如下表所示：

表：資料集基本資料

資料集	維度	類別	訓練集筆數	測試集筆數
Sonar	6	15	225	75
92AV3C9	9	16	10366	10366
Satimage	36	6	4435	2000
Forest	54	7	11340	565892

四、結果與討論

4.1 分類技術之比較

表：全部屬性之分類結果(%)

資料集	維度	DT	NBC	BPN	SVM	kNN(k)	Fuzzy-kNN(k)
Sonar	6	46.67	90.67	78.67	<u>98.66</u>	<u>98.66</u> (3)	97.33(3)
92AV3C9	9	60.80	49.36	48.67	<u>93.63</u>	91.62(3)	36.68(3)
Satimage	36	81.35	78.6	84.15	<u>91.40</u>	90.65(3)	61.2(25)
Forest	54	46.73	33.87	42.21	<u>58.00</u>	N.A.	N.A.

上表結果顯示，就分類技術的比較上 SVM 分類法均比其他分類法有更優秀的表現，kNN 分類法居次。在 Sonar 資料集上 SVM 與 kNN 分類法均可得到最高得準確度 98.66%，SVM 屬於監督式學習法，故需有訓練階段，而 kNN 屬於案例式學習法，不需訓練模式，但其範例集不宜過大。

表：Satimage 屬性篩選之分類結果(%)

維度	DT	NBC	BPN	SVM	kNN(k)	Fuzzy-kNN(k)
36	81.35	78.60	84.15	<u>91.40</u>	90.65(3)	61.20(25)
24	80.20	79.05	84.80	90.70	90.30(5)	61.75(15)
12	80.20	80.55	84.35	89.65	88.85(10)	59.70(15)
4	74.60	78.25	82.05	84.45	84.35(5)	51.95(25)
中心 pixel (17,18,19,20)	50.40	80.70	80.45	85.55	85.10(25)	59.60(25)

Satimage 資料集中每筆資料由九宮格之 9 個 pixel 組成，共含有 36 個屬性(= 4 spectral bands × 9 pixels in neighborhood)，每個屬性為介於 0-255 的數值。在屬性篩選的比較上，本研究分別使用 36 個、24 個、12 個、4 個、中心 pixel 的部分屬性進行分類準確度的比較，其中中心 pixel 係採取中間 pixel 的四個屬性(17, 18, 19, 20)來代表。研究結果顯示大體上屬性越少，其準確度也逐漸降低，但所需的演算時間也跟著降低。但是在屬性由 36 個降低至 24 個時，以 SVM 為例準確度僅降低 0.7%，但時間節省僅為原來的 67%。

4.2 分群技術之比較

表：正規化之分群結果(%)

方法	Fuzzy C-means			ISODATA			分群數
	原始	M 正規	Z 正規	原始	M 正規	Z 正規	
Sonar	85.33	<u>92.00</u>	90.67	84.00	<u>92.00</u>	90.67	16
92AV3C9	49.85	51.72	52.42	49.46	51.63	<u>52.83</u>	15
Satimage	71.55	72.30	72.20	71.55	<u>73.75</u>	72.20	7

以 Sonar 資料集為例，經 M 正規化後，不管以 Fuzzy C-means 或 ISODATA 分群法，均能分出最高的準確度為 92%。92AV3C9 資料集經 Z 正規化後以 ISODATA 分群法，可分出最高的準確度為 52.83%。Satimage 資料集經 M 正規化後以 ISODATA 分群法，可分出最高的準確度為 73.35%。

表：維度縮減之分群結果(%)

資料集	準確率(維度)			分群數
Sonar	85.33(6)	<u>97.33</u> (5)	<u>97.33</u> (3)	15
92AV3C9	49.85(9)	<u>52.81</u> (5)	47.39(3)	16
Satimage	<u>71.55</u> (36)	69.95(5)	70.35(3)	7

在 Sonar 資料集方面維度由原來的 6 為縮減為 3 及 5 維後，準確度從 85.33 提高至 97.33。在 92AV3C9 資料集方面維度由原來的 9 縮減為 5 維後，準確度從 49.85 提高至 52.81。但在 Satimage 方面降低了維度，準確度卻也跟著降低。

表：群數不同之分群結果(%)

資料集	維度	準確率(群數)		
Sonar	3	97.33(15)	97.33(19)	<u>100.00</u> (22)
92AV3C9	5	52.81(16)	54.04(24)	<u>54.97</u> (32)
Satimage	3	70.35(07)	<u>77.75</u> (11)	76.80(14)

在 Sonar 資料集方面將維度先固定為 3 維，然後再改變分群群數後比較，由原來的 15 群為改變為 22 群後，準確度從 97.33 提高至 100。在 92AV3C9 資料集方面將維度先固定為 5 維，然後再改變分群群數後比較，由原來的 16 群為改變為 32 群後準確度從 52.81 提高至 54.97。在 Satimage 資料集方面將維度先固定為 3 維，然後再改變分群群數後比較，由原來的 7 群為改變為 11 群後，準確度從 70.35 提高至 77.75。

4.3 分類與分群技術之綜合比較

表：綜合比較結果(%)

資料集	分群法	DT	NBC	BPN	SVM
Sonar	<u>100.00</u>	46.67	90.67	78.67	98.66
92AV3C9	54.97	60.8	49.36	48.67	<u>93.63</u>
Satimage	77.75	81.35	78.6	84.15	<u>91.40</u>

以 Sonar 資料集為例，分群法之準確度為 100% 優於決策樹、簡單貝式、類神經等分類演算法。但以 92AV3C9 資料集為例，分類演算法中的 SVM 法之準確度為 93.63%，比其他的分類演算法及分群法都有更高的準確度。而在 Satimage 資料集方面，分類演算法中的 SVM 法之準確度為 91.40%，比其他的分類演算法及分群法都有更高的準確度。

五、結論

本研究使用五個資料集來測試比較五個多高譜的衛星影像資料集，探討決策數(DT)、單純貝氏分類(NBC)、類神經網路(BPN)、支持向量機(SVM)、最近芳鄰法(kNN)、模糊最近芳鄰法(Fuzzy-kNN)等六種分類演算法；及K-Means法、Fuzzy C-Means法、ISODATA法等三種分群演算法對於衛星影像資料集之多光譜影像判識效能進行比較，可得知應用現代資訊科技發展自動衛星影像分類器機制將比傳統人工分類更可對使用者提供一個更快速方便的分類服務。

實驗結果指出，在分類準確性方面：在所有資料集的比較結果均以SVM法能有最高的分類準確度，而kNN分類法的分類準確度表現上也優於SVM以外的其他分類法。以Sonar資料集為例，以SVM的分類準確度98.66%最高，以決策樹46.67%最低。以92AV3C9資料集為例，以SVM的分類準確度93.63%最高，以Fuzzy-kNN 36.68%最低。以Satimage資料集為例，以SVM的分類準確度91.4%最高，以Fuzzy-kNN 61.2%最低。以Forest資料集為例，以SVM的分類準確度58.00%最高，以單純貝氏分類33.87%最低。

實驗結果指出，在分類執行速率方面：以kNN分類法為最快，其次為Fuzzy-kNN、DT、NBC、BPN，最慢的為SVM法。因SVM法需一段時間來分析分類最佳參數，所以所需時間較長。但也因依據最佳參數進行分類，故分類準確性也最高。

實驗結果指出，大體而言，SVM分類法比其他的分類演算法及分群法都有較高的準確度表現。

六、參考文獻

因篇幅限制，本參考文獻省略。