

# 行政院國家科學委員會補助專題研究計畫成果報告

※※※※※※※※※※※※※※※※※※※※※※※※※※※※※※

※

※

※

地理資訊系統中資料發掘之研究

※

※

Data Mining in Geographic Information Systems

※

※

※

※※※※※※※※※※※※※※※※※※※※※※※※※※※※※

計畫類別：個別型計畫      整合型計畫

計畫編號：NSC 89-2218-E-002-025

執行期間： 89年 08月 01日至 90年 07月 31日

計畫主持人：歐陽彥正

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

執行單位：國立台灣大學資訊工程學系

中華民國 91 年 01 月 11 日

# 行政院國家科學委員會專題研究計畫成果報告

## 地理資訊系統中資料發掘之研究

### Data Mining in Geographic Information Systems

計畫編號：NSC 89-2218-E-002-025

執行期限：89年08月01日至90年07月31日

主持人：歐陽彥正 國立台灣大學資訊工程學系

計畫參與人員：葉建華 國立台灣大學資訊工程學系

#### 一、中文摘要

本計畫的目標是研究一套能自動分析及歸納不同年代地圖中，空間物件變遷關係的演算法，並將變遷關係以一連串基本變遷因子，如平移、旋轉、尺寸改變量加以描述。同時亦針對形狀改變，研究出一套定量的測度，其目的是針對儲存於數位圖書館中的大量向量地圖，以所研究出的演算法加以處理。而後以人類思考能輕易掌握的描述方式，歸納結論出不同版本地圖間，空間物件的變遷關係。因此本研究可以說是針對二度空間的資料進行 data mining。

本研究最重要的課題是針對空間中兩個幾何圖形的演變關係，將之拆解為人類思考能輕易掌握的描述方式，使其輸出具有知識價值的資訊，而非只是一連串難以了解其背後涵意的數字。

本計畫的成果除了在資訊領域上會有學理的貢獻之外，並能有效運用在地理資訊系統的內容分析上，對追蹤國土改變、土地使用及建立文化地圖上均有相當大的應用價值。

**關鍵詞**：data mining, geographic information system

#### Abstract

The main purpose of this project is to develop a data mining algorithm for mining evolutionary changes of spatial objects on digitized maps. The data mining algorithm should summarize the evolutionary changes in a way that can be easily interpreted rather than just provide a set of numbers that does not fit our intuition. The main challenge of

this project is that if the evolutionary changes are to be decomposed into a sequence of atomic operations such as translation, rotation, and scaling, the decomposition is generally not unique. Therefore, the issue of what decomposition is optimal arises. In this regard, we need to define an optimization object function and figure out an optimization mechanism. Another main challenge is how we can summarize the change of shape in a quantitative way.

The research results of this project not only will contribute to advances of data mining research but also have significant real world applications, including urban planning and cultural atlas.

**Keywords**: data mining, geographic information system

#### 二、緣由與目的

建構數位圖書館並將大量各種型態的資料數位化是近年來資訊化社會最重要的發展趨勢之一，然而在大量資料數位化後，首先必須思考的問題便是如何有效的運用這些資料。由於資料量龐大，欲有效運用資料便必須藉助電腦的處理能力先將資料進行分析或發掘其中隱含的知識。因此 data mining 及相關的研究便成為近幾年來熱門的研究議題〔1〕〔2〕。

本計畫的主題，是研究一套分析地圖上空間物件變遷的演算法，藉以自動偵測不同年代空間物件之間的變遷關係，並以人類思考所能輕易掌握的方式加以歸納，所以可說是針對二度空間資料的 data mining。如前所述，在目前數位圖書館的快速發展下，資料大量數位化已經蔚成潮

流，因此地圖的數位化也成為數位圖書館中的主要工作之一。在這樣的環境下，要找出不同版本地圖之間地理物件的變遷關係，若仍以人工方式進行，將會耗費相當多的人力與時間。有鑑於此，如何開發的一個有效且快速的自動變遷偵測方法，將有助於地理資訊方面的研究與應用。

如前所述，由於大量資料的數位化 Data mining 已成為近年來資訊領域的熱門研究議題。然而針對空間物件方面的 data mining 仍不多見。因此本計畫在這方面具有相當的指標意義。至於在應用方面，本研究的成果將可以應用於分析城鎮的變遷、土地利用變化、河道改變等，對於國土的規劃與利用相當有幫助。如前所述，目前地理資訊系統已有利用衛星影像探討地形地貌的改變。本研究的成果將可以與利用衛星影像的方法相輔相成，大量應用在追蹤國土改變、土地使用狀況，以及文化地圖的建立等方面。

### 三、結果與討論

#### 定義空間物件變遷比對

在分析空間物件的變遷上，首先必須面對如何在不同年代的地圖中，找出兩兩相對應的物件。當然不同版本地圖上的空間物件因為年代變遷或測量方法不同的關係，其絕對位置與相對位置不可能一樣，因此在比對空間物件時就不能以其位置座標來當做量測的方法。在我們進行中的計畫是以 fuzzy logic 表現空間物件的座標關係，再據此計算物件間空間座標的相似度，而得到相當精準的效果。

本計畫的研究是以上述的成果為基礎進一步分析空間物件的變遷。至於所謂空間物件的變遷比對，從直觀上而言，就是將兩個相同的物件做比較，找出最可能的變遷型態，包含物件可能經過什麼樣子的位置移動，物件可能經過某些角度的轉動以及物件可能經過某種比例的縮小或放大等。

我們將物件的相似度定義如下：

當兩個物件經過平移、旋轉、放大/縮小，以求得面積正規化後，其重疊面積比率為最大時，則認為這兩個物件處於最相似的

情況。式子如下：

$$\text{Area Ratio} = (\text{Area of Object1} \cap \text{Object2}) / (\text{Area of Object1} \cup \text{Object2})$$

決定了最佳化目標函數之後，就可以開始準備尋找變遷的方式了。所謂的變遷方式其實就是最前面所述的座標轉換 M，當物件 A 經過一個座標轉換 M 之後與物件 B 最相似，則這個座標轉換 M 即為物件 A 與物件 B 的位相差。座標轉換 M 的組成如下：

$$M = T(\text{CentroidX}, \text{CentroidY}) \times T(dx, dy) \times R(\theta) \times S(Sx, Sy) \times T(-\text{CentroidX}, -\text{CentroidY})$$

或

$$M = \begin{bmatrix} 1 & 0 & \text{CentroidX} \\ 0 & 1 & \text{CentroidY} \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & dx \\ 0 & 1 & dy \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} Sx & 0 & 0 \\ 0 & Sy & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & -\text{CentroidX} \\ 0 & 1 & -\text{CentroidY} \\ 0 & 0 & 1 \end{bmatrix}$$

其中 T(dx, dy) 為平移運動矩陣，R(θ) 為旋轉運動矩陣，S(Sx, Sy) 為放大/縮小運動矩陣。

其中 CentroidX 及 CentroidY 為物件 A 的中心點，是一個常數，因此 M 其實是由五個自變數 (dx, dy, Sx, Sy, θ) 構成的一個矩陣，而本研究研發出一個 Optimization Algorithm，能在這五個自變數的搜尋空間裡儘可能找到一個最佳化的矩陣 M，使得兩個物件的相似度最大；換句話說，利用最佳化演算法調整五個變數，求取 F(M) = O/U 的極大值，使得兩個物件的重疊面積比率為最大，以求得面積比率的極限值。

### 四、計畫成果自評

本計畫的主要成果為研發出如何由兩個不同年代地圖中，比對空間物件間的變遷關係，並能將變遷過程以一連串基本變遷因子加以歸納描述。整個研究主要是針對向量式的地理資訊。本研究所提出的變

遷演算法完全根據人類觀察直覺的基本變遷因子，如平移量、旋轉量及尺寸改變量作為描述基礎。

至於具體應用，我們預期可以相當成功地將兩個不同地圖集裡的空間物件(如聚落、河川等)拿來比較，根據其絕對位置與外觀形狀，求得一個變遷方式的總分析與歸納。亦即往後人們對於地理資訊的處理，可以將相對簡單但繁瑣的資料分析及較為直接的歸納工作交由電腦處理即可，如此將能大量節省人力及時間。

此外，從空間物件變遷比對的整個理論來看，我們發現除了可用來比對城鎮聚落的變遷之外，未來尚可以在幾個大方向上面做一些研究：

1. 變遷因素的研究：我們目前的研究只著眼於變遷現象的抽取與描述，並沒能夠更進一步去研究造成空間物件變遷的原因，如若能將變遷結果與環境變遷及交通動線的改線等因素整合起來，並利用資料採礦(Data Mining)的技術，進一步去分析隱藏的變遷原因，則相信更能夠發掘出更多有趣的結論。如某個聚落的發展方向是否直接或間接受到交通動線的改變或河道變遷的影響等。
2. 展示變遷過程：使用視覺化的顯現效果(Visualization)，將本研究取出的變遷特徵值，模擬變遷過程，並將之顯象呈現在使用者面前，以方便使用者了解，並增加變遷比對的趣味性。

## 五、參考文獻

- [1] Chung-Chen Lai, Jian-Hua Yeh, and Yen-Jen Oyang. Toward Automatic Generation of Geographic Thesaurus through Data Mining in a Geographic Information System. In *Proceedings of The Second Asian Digital Library Conference*, pages 202--214, Taipei, Taiwan, R.O.C, November 1999.
- [2] Meng-Han Yang (2001), Spatial Objects Matching and Transition Analysis Applied in Data Mining, Department of Computer Science and Information

Engineering, National Taiwan University, 2001。

- [3] Chung-Chen Lai (1999), A Study on Matching Spatial Objects in Geographic Information Systems, Department of Computer Science and Information Engineering, National Taiwan University, 1999。