

# 知識天地

## 流形學習與人臉辨識

劉庭祿（資訊所副研究員）

在計算機科學的相關研究領域裡，常需利用向量的形式來描述所欲探討之資料物件。以視訊影像為例，一張解析度為 64 x 64 像素的灰階影像，可用一 4096 維度的特徵向量（feature vector）來表示，其中每一維度儲存了相對應像素的灰階值。將物件以特徵向量之數學化形式呈現後，一個物件即可被視為向量空間中的一點，而一群物件將會在空間中形成某種分佈，這有益於實際應用上的分析，如物件的特徵萃取（feature extraction）、分群（clustering）、分類（classification）或辨識（recognition）。

直覺上，若將資料物件表示的越詳盡，或提供更多的特徵於向量表示中，應可提昇問題處理上的正確率才對；然而在絕大多數實例中，這樣的做法是有其極限的。主要原因在於當採用較複雜物件特徵時，也相對提高了向量空間的維度，而在過高維度的空間裡，物件的樣本數需要以近乎指數的成長（exponential growth），才能維持穩定的資料分析正確率。由於實際應用上可用資料往往是有限的，如此一來將物件特徵表示的較為詳盡之好處，並無法彌補高維度空間所導致樣本數不足及其衍生問題，這個現象稱之為維度的詛咒（the curse of dimensionality）。另一方面，近來由於多媒體應用的蓬勃發展，常需要分析如視訊、聲音、超文件或高解析度醫學影像等資料，這些物件其特徵向量在本質上往往就是高維度的，為了計算上的效率與避免維度的詛咒，一般可於初始的物件表示上，使用較高維度的向量盡量去包含此類物件的特徵資訊，再透過降維（dimensionality reduction）演算法，來萃取或保持其在高維度空間裡所隱含的重要性質。降維分析可適用於資料分類或分群，亦可作為其他進階應用的前處理。而我們所要介紹的流形學習（manifold learning）就是一種降維分析的方法。

希臘哲學家赫拉克利特斯（Heraclitus）曾說：「一個人不可能踏入同一條河兩次」，因為萬物皆在流轉變動；以人臉來說也是如此，隨著各種不同的內在及外在變因，即使是同一個人的臉也會顯現出不同的外徵（appearance）。較嚴格來說，我們亦不曾看過同一張臉兩次，然有趣的是人類的視覺系統並不會因此感到困擾，而造成人臉辨識上的困難。不論是同一條河或同一個人，不管外徵型態有多不同，我們大都能輕易地加以辨識。Seung 與 Lee [3] 曾在 Science 期刊提出以下的觀點來解釋此一現象，假設使用影像的形式記錄人臉，並將影像向量化，則每一張人臉就成為影像空間中的一點，而影像空間的維度即是影像的像素數。若觀察平滑變動下的人臉影像，其在影像空間中，會沿著一個低維度（遠低於影像空間的實際維度）的連續曲面移動；連續的原因是基於平滑的變動，而低維度曲面的成因則是基於變動因素往往只為極少數。舉例來說，在影像空間中（同平面）旋轉的人臉其變動維度可能只有一維，即是攝影機的旋轉角度。在圖一中，顯示 2 組不同旋轉角度的人臉影像，由於變動的維度只有旋轉角度一項，在高頻

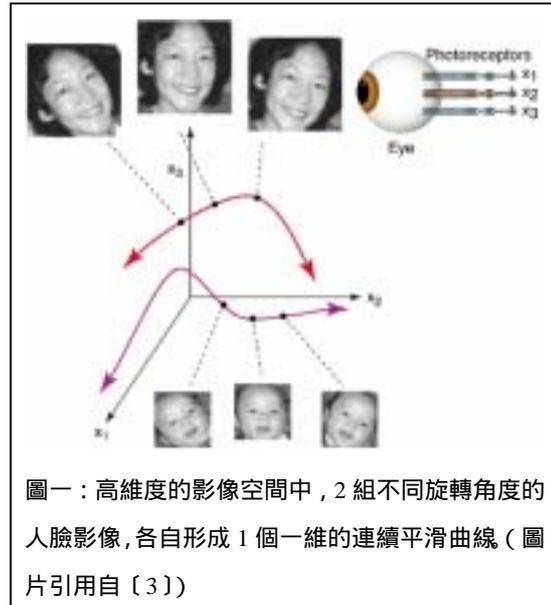
率拍攝情況下，2 組影像資料在（高維度）影像空間中，各自形成一維（低維度）的連續曲線。Seung 與 Lee 的所提及的（相較於原影像空間之）低維度曲面在數學上稱之為流形（manifold）。

（註：簡單來說，流形空間在定義上為一拓撲空間，其區域性（local）子空間可被視為為歐氏空間（Euclidean space），亦即其二相鄰座標點間之距離可用歐氏距離（Euclidean distance）來衡量。）

流形學習的方法就是基於所要分析的物件資料在高維度空間中有平滑流形

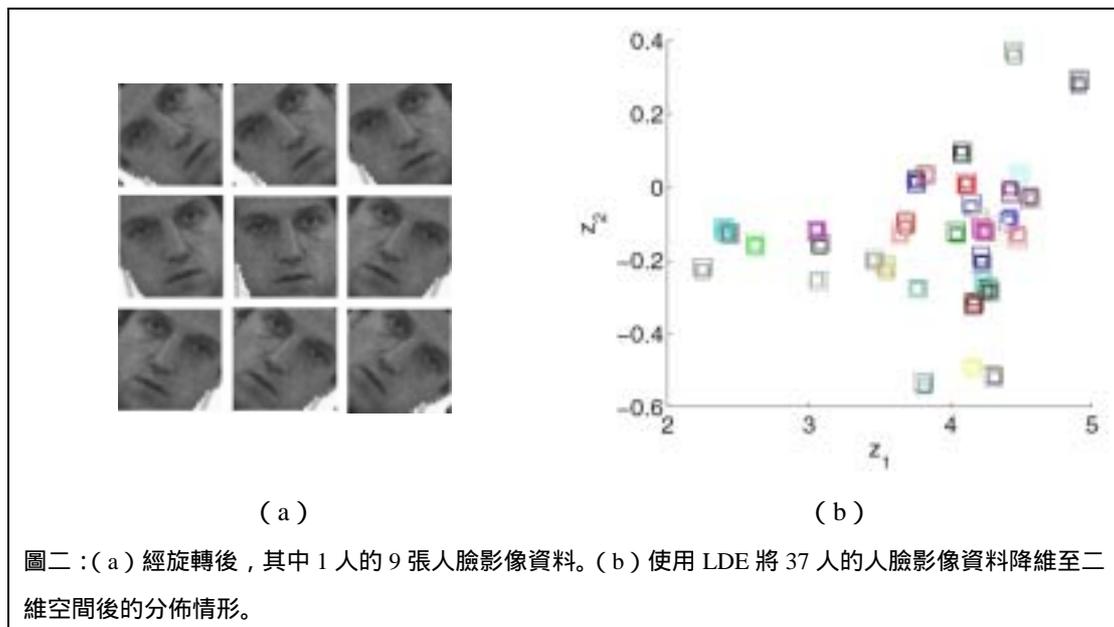
分佈的假設，再利用轉置（re-embed）的方法將物件資料映到較低維度的歐氏空間，並區域性（locally）保持其原有在流形上的分佈。在 2000 年的 Science 期刊中，有兩篇非常具貢獻性的流形學習演算法被提出，一是由 Tenenbaum 等人所發表的 Isomap (ISOmetric feature MAPping) [4]，另一是由 Roweis 與 Saul 所提出的 LLE (Locally Linear Embedding) [2]（註：這兩篇論文的作者今年都將出席中央研究院資訊所舉辦的 2006 年 Machine Learning Summer School）。透過 Isomap 與 LLE 的使用，可將原先分佈於高維空間的資料點，在低維度的空間中呈現，並保持這些資料點在高維度空間中的重要分佈結構；若降維後的空間，其維度設定為二維，則這些資料點將變為可視化（visualize），這對許多研究都是具有相當幫助的。

然而 Isomap 與 LLE 仍有其受限制之處，這兩個流形學習方法主要是透過降維對資料之可視化進行最佳化處理，著重於對訓練資料的表示（representation）而不是分類（classification）；且有 out of sample 的問題，亦即是當有新的物件資料時，就需要重新學習出轉置降維。有鑑於此，我們發展出一個取名為 Local Discriminant Embedding（簡稱 LDE）[1] 的監督式學習演算法，針對具有標示類別（class label）的物件資料，進行以分類為目標的最佳化處理。LDE 演算法的特色為建構於二個物件資料鄰近關係圖（neighborhood graph），分別用於記錄物件間的類別關係，與鄰近幾何關係，並可將最佳化問題推導至解特徵值問題，因此最佳降維空間可經由簡易的特徵值與特徵向量的計算而獲得。此外 LDE 沒有 out of sample 的問題，可直接處理新的測試資料。由於人臉資料的分佈在影像空間中常有流形的特性，且有較多的標準資料集可供測試，我們遂將 LDE 應用於人臉的辨識。由於篇幅所限，以下僅概述一個簡單的例子：係考慮 CMU 大學所提供的 PIE 人臉資料庫進行演算法測試，從其中挑選出 37 個未帶眼鏡者，如圖二（a）所示，將人臉影像作 $\pm 40$ 度的旋轉（每次增加 10 度），個別產生 9 張人臉影像；再使用 LDE 將這 37 人的影像資料（旋轉後共有 333 張）降維至二維



圖一：高維度的影像空間中，2 組不同旋轉角度的人臉影像，各自形成 1 個一維的連續平滑曲線（圖片引用自 [3]）

可視空間中，如圖二 (b) 所示。若每一張人臉影像以一正方形來表示，不同顏色用以區隔不同人，從圖二 (b) 中可看出經 LDE 降維轉置後，資料的分佈確實對辨識有最佳化的效果，亦即有同一類別的資料點聚集，而不同類別的資料點分開的性質。事實上，這個資料集經 LDE 降維至三維後，就可將來自這 37 個人的影像完全分對。



參考文獻：

- [1] H.-T. Chen, H.-W. Chang, and T.-L. Liu. Local Discriminant Embedding and Its Variants. In Proc. Int'l Conf. on Computer Vision and Pattern Recognition, volume 2, pages 846-853, 2005.
- [2] S. Roweis and L. Saul. Nonlinear Dimensionality Reduction by Locally Linear embedding. Science, volume 290, pages 2323-2326, 2000.
- [3] H. Seung and D. Lee. The Manifold Ways of Perception. Science, volume 290, pages 2268-2269, 2000.
- [4] J.B. Tenenbaum, V. de Silva and J. C. Langford. A Global Geometric Framework for Nonlinear Dimensionality Reduction. Science, volume 290, pages 2319-2323, 2000.