

尋優支持向量機於分類應用之研究

Applying Optimal Support Vector Machine for Pattern Classification

林錦昌

佛光人文社會學院 資訊學系

linedwin@ms14.hinet.net

駱至中

佛光人文社會學院 資訊學系

locc@mail.fgu.edu.tw

摘要

近十年來，支持向量機(Support Vector Machine，簡稱為 SVM)由於其在許多領域問題之解決上均有良好表現而受到矚目。在支持向量機中包含一個學習演算法及其輸入空間，輸入空間內含一個訓練集和一個測試集。藉由訓練集的輸入，學習演算法可以找出一個辨別器，並經由測試集我們可以知道此辨別器的辨別正確度。支持向量機的目的便是找出正確度夠好的辨別器以分辨鑑別後續未知的輸入項。機器學習機制學習效果的優劣，會因所使用的訓練資料集中能否涵蓋整個問題空間的而定。但隨著輸入空間的維度增加，計算的複雜度也跟著提高，而往往致使機器學習效果變差。特性篩選便是為解決這類問題而發展出來的。而特性篩選又可分為二種，一為針對輸入空間中的向量分量作排序，以決定分量對辨別器的影響，進而捨棄不重要者以降低計算的複雜度；其二則以找出含蓋整個問題空間的最大樣模，來做為訓練範例以得到最佳的學習機制。在本論文中，我們提出「尋優支持向量機」，其中以模擬退火演算法找出支持向量機所需的支持向量集合做為訓練資料集，配合支持向量機的學習機制，在較少數量訓練資料集中，找出最能代表整個母體資料的訓練樣本，建構最大邊緣解的支持向量集以克服大量的輸入資料的分類問題。為驗證所提出之機器學習機制的分類效能，本研究以糖尿病資料集為驗證實例。結果顯示，本研究所提出之尋優支持向量機確實有能力處理樣模分類問題。

關鍵詞：支持向量機(Support Vector Machine)、模擬退火演算法(Simulated Annealing)、尋優支持向量機(Optimal SVM)、特性篩選(Feature Selection)

一、緒論

支持向量機(SVM)源起於統計學習理論，它研究如何構造學習機制以解決樣模分類

(Pattern Classification)問題[8]。支援向量機使用結構風險最小化(Structural Risk Minimization, SRM 準則)原理構造決策超平面，使每一類資料之間的分類間隔(Margin)最大[11]。根據 SRM 準則：學習機對未知數據分類所產生的實際風險在 $0 < \eta < 1$ 時，得滿足以下關係：

$$R \leq R_{emp} + \sqrt{\frac{h(\log(2n/h) + 1) - \log(\eta/4)}{n}} \quad (1)$$

其中， n 為訓練樣本個數， R 是實際風險，而不等式的右邊總稱為「風險邊界」， R_{emp} 為經驗風險， $\sqrt{\frac{h(\log(2n/h) + 1) - \log(\eta/4)}{n}}$ 項則被稱之為“VC 置信值”，值項中的 h 是學習機的 VC 維，它也反映了此學習機的複雜程度。

SVM 的基本精神就是在樣本數目適宜的前提下，選取比較好的 VC 維(h)，使 R_{emp} 和 VC 置信值得以折衷均衡，使每一類別資料之間的分類間隔(Margin)最大，最終使得 R 值變小。對於線性不可分的數據，支援向量機則依照 Cover 定理，通過核函數把資料由低維空間向高維空間映射，進而在高維空間為低維資料構造線性可分超平面。SVM 研究中，最常用的核函數主要有多項式函數、徑向基函數和多層 Sigmoidal 神經網路函數等三類。而其空間轉換的具體作法為：在低維不可分的情況下，對於每個要判別的未知樣本 u ，計算高維空間中的最佳超平面分類函數 $f(u)$ ：

$$f(u) = \sum_{i=1}^S \alpha_i y_i K(u, x_i) + b \quad (2)$$

其中， α_i 是支持向量 x_i 對應的 Lagrange 乘數， y_i 是支持向量 x_i 的類別標記(值為 1 或者 -1)， S 是支持向量總數， K 是核函數， $f(u)$ 值的正負即分別代表不同的類別。支援向量機適合解決高維度的稀疏矩陣資料，但當要克服大量的輸入資料，找出支持向量矩陣時，將需要極大

的資源及複雜的計算時間。

模擬退火演算法(Simulate Algorithm, 簡稱 SA)是屬於啟發式演算法的一種。而所謂的啟發式演算法, 均會有下列四個特點[1, 7]:

- 1、 在解空間中反覆的搜尋。
- 2、 問題中的決策變數扮演某些特定角色, 例如: 神經元(neurons)、微粒(particles)、染色體(chromosomes)、螞蟻(ants)等。
- 3、 決策變數間採競爭與合作的運算方法。
- 4、 演算法本身擁有對問題的表現, 進行自我修正改善(self-modification)的機制。

由以上說明可知, 啟發式演算法乃是一群由人工智慧所衍生或與之相關的尋優技術。它藉由決策變數的角色扮演, 在解空間中依循一定的法則進行搜尋, 而 SA 則是近年來公認最具有實用價值的啟發式演算法之一。一般在處理最佳化的問題時, 必須注意搜尋是否落入區域的最佳解, 在近年來所謂的啟發式演算法, 利用隨機性的搜尋法則, 在可行解空間中利用疊代改善的搜尋方法搜尋最佳解, 以提高得到全域最佳解(global optima)的機率。所謂的疊代改善法, 就是指演算法在搜尋的過程中, 在經過『初始解』的設定之後, 演算法會在鄰近的解空間當中, 尋找一個較佳的鄰近解, 並且將較佳的鄰近解, 取代而成為目前解, 如此經由不斷的疊代改善之後, 搜尋便會朝最佳解邁進, 找到最佳值。SA 便其中的一種相當成功的搜尋方法。

本文所提出一名為「尋優支持向量機」的機器學習機制, 機制中以模擬退火演算法配合支持向量機, 在數量較少的訓練資料集中, 找出最能代表整個母體空間的訓練樣本, 建構最大邊緣解的支持向量集以克服大量輸入資料的分類問題。本研究中並以糖尿病資料集為實證案例來驗證本研究提出之尋優支持向量機的效能及處理能力。

二、模擬退火演算法

1953 年, Metropolis 等人首先提出一套模擬金屬退火過程的電腦演算程序, 稱作 Metropolis 程序。之後, Kirpatrick 等人(1983)便以此作為理論基礎, 提出了 SA。所謂的 SA, 其原理乃是模擬凝結物質結晶的退火程序(annealing process)。其主要的構想在於當物體受熱, 溫度升高至熔融狀態時, 若將分子不同的結晶狀態模擬成不同的可行解, 在溫度高時, 分子的活動能力較強, 在解空間之中散亂的搜尋, 所形成結晶的型態較多, 也代表其搜尋範圍較大[2]。

理論上, 隨著退火程序的進行, 溫度隨著時間而下降, 物體分子的活動能力也跟著降低, 搜尋範圍減小, 結晶趨穩定, 當溫度降到最低時, 物體的結晶為最穩定的狀態, 我們稱之為基態 (ground state), 代表其分子間的最小能量, 且晶格的排列為最佳的狀態, 圖 1 為模擬退火演算法搜尋過程的演化歷程圖。然而, 在溫度降低的過程中, 若此降溫過程太快, 亦即快速地使熔融狀態的物體達到固化狀態, 物體快速地凝結會使得部份晶格來不及重新排列, 如此材料內部就會有晶格排列紊亂的情形發生, 這一種狀態的能量將會比經過充分時間退火的材料的能量高, 這種情形稱之為快速淬火 (rapid quenching)。快速淬火的結果, 導致晶格排列無法達到最佳的狀態, 亦即函數落入了區域的最佳值。此時必須利用另一個方法, 用機率公式來決定協助 SA 跳脫區域的最佳解, 這個方法稱之為波茲曼機率分布 (Boltzmann probability distribution)。

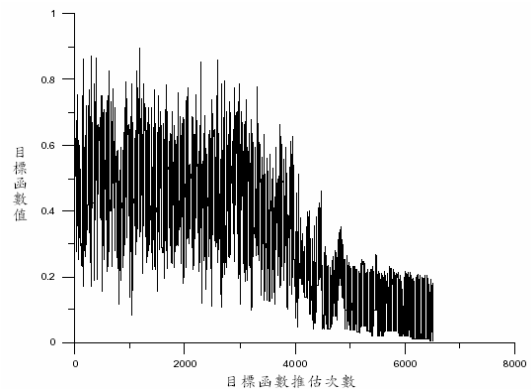


圖 1、模擬退火演算法搜尋過程

SA 屬於疊代改善法的一種, 除了疊代改善之外, SA 之所以能夠成功的搜尋全域最佳解, 在於它改善了疊代改善法容易陷入區域最佳解的缺點, 而擁有跳脫局部最佳解的能力之故。所謂的解, 指的就是一種狀態, 以最佳化問題為例, 就是每個決策變數所對應的目標函數值, 就稱為一個『可行解(solution)』, 目前的可行解狀態就叫做『現行解』。以最小化的問題為例, 疊代改善法之『現行解』若是落在區域最小值的區域內, 便會繼續的往該區域之最小值移動, 而忽略了真正的全區域最小值。以下為模擬退火演算法 SA 的演算步驟[3, 4]:

- Step1、 設定初始狀態 $T = T_i$, 在空間中任意選取一點 S 作為起始點, 作為現行解, 並計算其目標函數值 $C(S)$ 。
- Step2、 利用某一個有效的擾動機制 (Perturbation Mechanism), 根據現行解的狀態, 產生一鄰近解 S' 。
- Step3、 計算 S 與 S' 目標函數值的差,

$$\Delta E = C(S') - C(S)。$$

- Step4、進行如下之決策判斷：若是 $\Delta E < 0$ ，此時表示鄰近解的狀態較佳，則接受 S' 成為新的現行解。反之，若是 $\Delta E \geq 0$ ，則以機率值 $p = \exp\left(\frac{-\Delta E}{KT}\right)$ 來判斷是否接受鄰近解；其作法為隨機產生一介於 0 與 1 之間的實數 R ，若 $p \geq R$ 則接受，反之則否。
- Step5、判斷收斂條件(終止條件)，若符合之前所設定的終止條件，則停止演算法，以目前解為最佳解，若否，則進行 Step6。
- Step6、降低溫度，利用溫度降低的機制，產生下一個溫度 $T_{i+1} < T_i$ ，重複 Step2，進行下一個搜尋。

三、支持向量機

支持向量機(SVM)是一個具有良好學習機制的特性，主要作用在產生一個用來分類的超平面(hyper plane)，並嚐試之間的空間最大化。更詳細的說 SVM 是一個用風險最小化的學習機制實作，其主要的原理在於用訓練錯誤率(training error rate)限制發生於 VC 維度之測試資料集的錯誤率。

通常，支持向量學習演算法 (SV algorithm) 可以被用來建構 (1)多項式學習機制 (Polynomial learning machines)、(2)半徑基底函數類神經網路(Radial-basis-function networks)和(3)多層式感知機學習機制(Multilayer perceptrons)等三種不同的學習機制。以下將就支持向量機的應用模式及理論進一步的闡述 [6, 7, 9, 10]。

3.1 支持向量機於線性可分割 (linear separable patterns)的處理模式

在線性可分割的模式，當以 W 代表 weight vector, X 是 input vector, b 是 bias, d_i 是 output 則：

$$\begin{aligned} W^T X_i + b &\geq 0 & \text{for } d_i &= +1 \\ W^T X_i + b &< 0 & \text{for } d_i &= -1 \end{aligned} \quad (3)$$

上式(3)中的 W 及 b 可做調整而為值大於等於 1 或小於-1，Support vector 是那些合併在決策面的點，因此是最難合併的點。故可利用二次最佳化找出最佳的決策邊界超平面 (Quadratic optimization finding the optimal hyperplane)

在下式(4)中找出 w 及 b 的最佳解，以滿足其限制條件。

$$d_i(w^T x_i + b) \geq 1 \quad \text{for } i=1,2,\dots,N \quad (4)$$

且 w 符合以下式(5)之狀況：

$$\text{minimizes } \Phi(w) = \frac{1}{2} w^T w \quad (5)$$

這種叫做 primal problem，其有下列特性：

- 1、 $\Phi(w)$ 是 w 的 convex 函數
- 2、 w 是最佳化過程之線性 Lagrange 乘數

$$J(w, b, \alpha) = \frac{1}{2} w^T w - \sum_{i=1}^N \alpha_i [d_i(w^T x_i + b) - 1] \quad (6)$$

接著，再利用 duality theorem 而可得到：

$$Q(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j d_i d_j x_i^T x_j \quad (7)$$

而其間各參數變化必得在下列限制之內：

$$\sum_{i=1}^N \alpha_i d_i = 0 \quad \& \quad \alpha_i \geq 0 \quad \text{for } i=1,2,\dots,N \quad (8)$$

$$w_0 = \sum_{i=1}^N \alpha_{0,i} d_i x_i \quad (9)$$

$$b_0 = 1 - w_0^T x^{(s)} \quad \text{for } d^{(s)} = 1 \quad (10)$$

3.2 支持向量機於不可分離 (Non-separable pattern) 的處理模式

當在線性不可分割的模式，目的在於建立對所有的測試資料集平均最小的分類錯誤。如有違反

$$d_i(W^T X_i + b) \geq +1, \quad i=1,2,\dots,N \quad (11)$$

則稱之為 soft，違反的情形有兩種分在正確區但在 region 內或分錯區。Hyper plane 為：

$$d_i(W^T X_i + b) \geq 1 - \xi_i, \quad i=1,2,\dots,N \quad (12)$$

如果求出來的 $\xi > 1$ 代表分錯區 $0 \leq \xi \leq 1$ 代表分對區。

最小化分錯的個數：

$$\Phi(\xi) = \sum_{i=1}^N I(\xi_i - 1) \quad (13)$$

$$\Phi(w, \xi) = \frac{1}{2} w^T w + C \sum_{i=1}^N \xi_i \quad (14)$$

C 是由機制的複雜度及分錯的個數決定，可由 (1)經驗、(2) 用 VC 來分析，結果如可分割 (separable) 模式的一樣限制，在 hyper plane 只差最後用 VC 來分析， $0 \leq \alpha_i \leq C$ ， C 是由使用者決定， b_0 最好是由 data point 中找出平均值。

再經由對偶理論(Duality theorem)找出最佳解

- 1、如果 primal problem 有最佳解，則 dual

問題也有最佳解，且最佳解還相同

- 2、為了 w_0 是 primal 的最佳解 α_0 是 dual 的最佳解且 w_0 必需滿足問題及下式

$$\Phi(w_0) = J(w_0, b_0, \alpha_0) = \min_w J(w_0, b_0, \alpha_0) \quad (15)$$

最後建構 SVM 的樣式辨認

- 1、把輸入向量非線性對映到高維度的特徵空間，把輸入及輸出隱藏住。
- 2、把對映出來的特徵空間(feature space)做分類。

3.3 最佳化設計支持向量機

不同的輸入有不同的非線性轉換 (non-linear transformation) 故有不同的 $\phi_i(x)$ 經由運算我們設 $k(x, x_i) = \phi^T(x)\phi(x_i)$ 結果為

$$\sum_{i=1}^N \alpha_i d_i K(x, x_i) = 0 \quad (16)$$

Mercer's theorem :

$$k(x, x_i) = \sum_{i=1}^{\infty} \lambda_i \phi_i(x) \phi_i(x_i) \quad (17)$$

若且唯若

$$\int_b^a \int_b^a k(x, x') \Psi(x) \Psi(x') dx dx' \geq 0, \quad \int_b^a \phi^2(x) dx < \infty$$

得到當 $\lambda_i \neq 1$ 第 i 個 image $\sqrt{\lambda_i} \Phi_i(x)$ 包含在一個擴張性特徵函數(eigenfunctions)之輸入向量 x 的特徵空間內，且這樣的特徵空間可能是無限大。

最佳化支持向量機：

$$Q(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j d_i d_j K(x_i, x_j) \quad (18)$$

滿足下列條件：

$$(1) \sum_{i=1}^N \alpha_i d_i = 0 \quad (2) 0 \leq \alpha_i \leq C \text{ for } i = 1.2 \dots N$$

$$(2) w_0 = \sum_{i=1}^N \alpha_{0,i} d_i \phi(x_i)$$

支持向量機的三種內積核函數：

$$(a) \text{ 多項式學習機制: } (x^T x_i + 1)^p \quad (19)$$

$$(b) \text{ RBF: } \exp\left(-\frac{1}{2\sigma^2} \|x - x_i\|^2\right) \quad (20)$$

$$(c) \text{ 二層式感知器: } \tanh(\beta x^T x_i + \beta_1) \quad (21)$$

SVM 解分類問題較均方差 (mean square error) 更合適，且不需要 domain knowledge，在非線性回歸，用 SV 方式比 MLP，更有效，計算更快。

3.4 SVM 解決非線性回歸問題的步驟

SVM 解決非線性回歸問題所使用的 loss function 如下式：

$$L_\epsilon(d, y) = \begin{cases} |d - y| - \epsilon, & \text{for } |d - y| \geq \epsilon \end{cases} \quad (25)$$

此式叫 E-insensitive loss function，其中 d 為想要的輸出， y 為評估的結果， ϵ 為規定的參數。其 Lagrange 方程式如下式：

$$y = \sum_{j=0}^{m_1} w_j \phi_j(x), R_{map} = \frac{1}{N} \sum_{i=1}^N L_\epsilon(d_i, y_i) \quad (26)$$

滿足下列條件：

$$d_i - w^T \phi(x_i) \leq \epsilon + \xi_i, w^T \phi(x_i) - d_i \leq \epsilon + \xi_i', \xi \geq 0, \xi' \leq 0, \text{ for } i = 1.2 \dots N$$

其最小成本函數：

$$\Phi(w, \xi_i, \xi_i') = C \left(\sum_{i=1}^N (\xi_i + \xi_i') \right) + \frac{1}{2} w^T w \quad (28)$$

$$J(w, \xi, \xi', \alpha, \alpha', \gamma, \gamma') = C \sum_{i=1}^N (\xi_i + \xi_i') + \frac{1}{2} w^T w - \sum_{i=1}^N \alpha_i [w^T \phi(x_i) - d_i + \epsilon + \xi] - \sum_{i=1}^N \alpha_i' [d_i - w^T \phi(x_i) + \epsilon + \xi'] - \sum_{i=1}^N (\gamma_i \xi_i + \gamma_i' \xi_i') \quad (29)$$

取得 convex function :

$$Q(\alpha_i, \alpha_i') = \sum_{i=1}^N d_i (\alpha_i - \alpha_i') - \epsilon \sum_{i=1}^N (\alpha_i - \alpha_i') - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (\alpha_i - \alpha_i') (\alpha_j - \alpha_j') K(x_i, x_j) \quad (30)$$

且須滿足下列限制條件：

$$\sum_{i=1}^N (\alpha_i - \alpha_i') = 0, 0 \leq \alpha_i \leq C,$$

$$i = 1.2 \dots N, 0 \leq \alpha_i' \leq C, i = 1.2 \dots N$$

四、研究設計

範例學習機制的資料量對其再生能力有重要影響。低資料量學習系統只需要較小的訓練集，高資料量學習系統則需要較大的訓練集，但其所獲得的解將優於前者。對給定訓練集來說，高資料量學習系統的訓練集誤差和測試集誤差之間的差別將大於低資料量學習系統。故如何找出具有涵蓋整個資料範例特徵的最小訓練資料集，將是克服範例學習人工智慧技術設計的一大瓶頸。神經計算技術已經在很多領域得到了成功的應用，但由於缺少一個統一的理論框架，經驗性成分相當高。這使得研究者們難以對各種神經計算模型的性能及其適用範圍進行理論分析，僅能用不十分可靠的實驗性比較評價優劣。另一方面，在利用神經計算解決問題時，也只能採取具體問題具體分析的

方式，通過大量費力耗時的實驗摸索，確定出合適的神經網路模型、演算法以及參數設置。這些缺陷已經對神經計算的進一步發展造成了極大的阻礙。

Vapnik 等人提出支援向量機制。由於其出色的學習性能，該技術已成為機器學習界的研究熱點。支援向量機是 Vapnik 等人提出的一類新型機器學習方法。由於其出色的學習性能，該技術已成為機器學習界的研究熱點，並在很多領域都得到了成功的應用，如人臉檢測、手寫體數位識別、文本自動分類... 等。支援向量機 (SVM) 是一種建立在統計學習理論基礎上的機器學習方法。通過學習演算法，SVM 可以自動尋找那些對分類有較好區分能力的支援向量，由此構造出的分類器可以最大化類與類的間隔，因而有較好的推廣性能和較高的分類準確率。SVM 主要思想是針對兩類分類問題，在高維空間中尋找一個超平面作為兩類的分割，以保證最小的分類錯誤率。而且 SVM 一個重要的優點是可以處理線性不可分的情況。用 SVM 實現分類，首先要從原始空間中抽取特徵，將原始空間中的樣本映射為高維特徵空間中的一個向量，以解決原始空間中線性不可分問題。

為了解決高維度的分類問題，本文中提出一個兩階段式的學習方法，底層以支援向量機 (SVM)，而上層以模擬退火演算法搜尋最佳的支持向量空間及支援向量機的參數組合。尋優支援向量機制利用 SA 搜尋具有涵蓋整個資料範例特徵的最小訓練資料集，以克服範例學習人工智慧技術設計的障礙。利用 SVM 自動尋找那些對分類有較好區分能力的支援向量，由此構造出的分類器可以最大化類與類的間隔，使其具有較好的推廣性能和較高的分類準確率來實現分類。

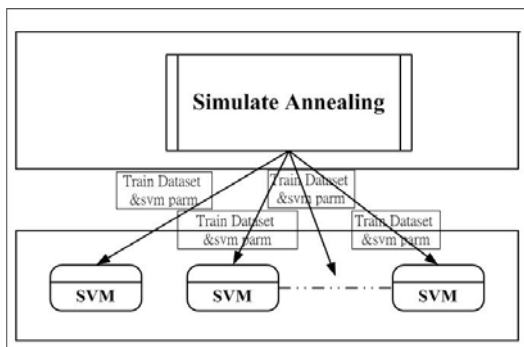


圖 2、尋優支援向量機模式 (SA-SVM)

本模式利用 SA 演算法在有限的輸入空間樣本資料量搜尋最能函蓋整個母體資料集的訓練資料集及搜尋最佳各支向量機制的調整參數，本模式的方法大致如下，圖 3 為本研究的模擬退火法之流程圖：

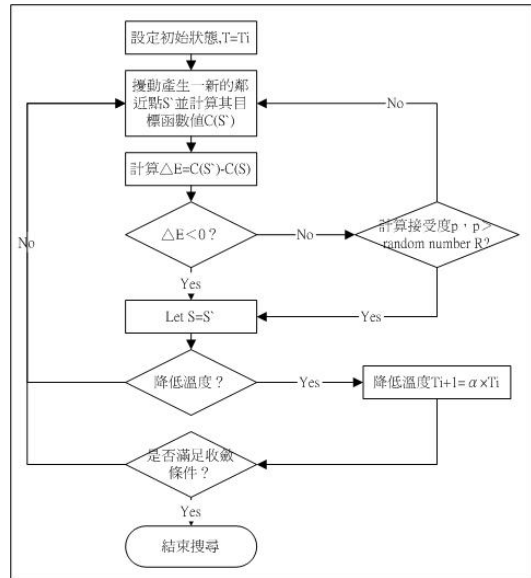


圖 3、SA-SVM 之模擬退火流程

- (a) 在可允許的參數空間內任取一亂數向量 x_1 ，並求其適應函數值 $f(x_1)$ 。函數值 $f(x_1)$ 為經由支援向量機訓練後的辨識正確率。
- (b) 給一個初始高溫 $T(0) = T_{init}$ 。
- (c) 用隨機或其他智慧型策略找出第二組解答向量 (solution vector) x_2 並求取 $f(x_2)$ 。
- (d) 求上述兩適應函數之差， $\Delta f = f(x_2) - f(x_1)$ 。
- (e) 如果 $\Delta f < 0$ ，表示能量在遞減中，則 x_2 的答案可以接受，跳至步驟 (f)。如果 $\Delta f > 0$ ，先求一概率 $p(k) = \exp[-\Delta f / T(k)]$ ，然後由亂數我出一值，若此值比 $p(k)$ 值小則 x_2 可接受，此表示金屬冷卻過程中有可能暫存在某一比較高位的能階上，再跳至步驟 (f)。如果上述概率不成立則至步驟 (g)。k 為進行次數的指標 (Iteration index)。
- (f) 令 $x_1 = x_2$ ， $f(x_1) = f(x_2)$ ，此表示把最好的結果存起來。其次降低溫度，依 $T(k+1) = \lambda T(k)$ 公式來算， λ 是一個 0 至 1 的實數，一般選用 0.95 至 0.98。 λ 的選定是依隨機取亂數 0.95 至 0.98 之間。

如果現在的溫度已比設定的最後低溫為小， $T(k) < T_{final}$ ，那最後結果為 x_1 及 $f(x_1)$ 。如果不成立則同步驟 (c)。

五、實例驗證與分析

本研究採用的實例驗證資料集為糖尿病 (Diabetes) 標準資料集為糖尿病檢測資料集，以個案生活習慣及相關檢驗值為輸入變數，計有 8 個輸入變數來判斷是否患有糖尿病，資料筆數為 768 筆，而其中 65.1% 不為糖尿病。資料來源：<http://www.liacc.up.pt/ML/statlog/datasets/diabetes/>。在本研究中，對標準資料集所設計的實驗流程如圖 4 所示，本研究將實驗資料模型劃分為三種資料集(測試資料集，佔 50%、驗證資料集，佔 25%、測試資料集，佔 25%)，系統是以 Matlab 語言為開發工具，在 Windows 平台上進行開發及實驗。

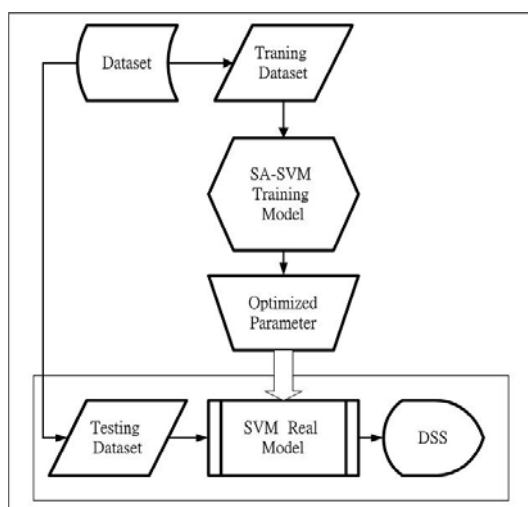


圖 4、本研究實例驗證實驗流程

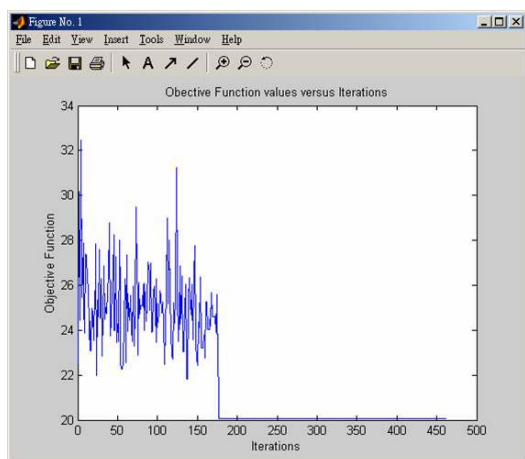


圖 5、SA-SVM 之模擬退火搜尋過程(linear)

本實驗首先底層以不同的支持向量機的核函數，線性核函數及半徑基底核函數，配合支持向量機的 langrange 學習演算法，找出分類決策超平面的極大化的分類間隔 (Margin)；再搭配上層模擬退火演算法，搜尋最小的訓練集，且具有涵蓋最大的分類決策

超平面的極大化的分類間隔 (Margin) 的支持向量，進而降低測試資料集中分類錯誤率。圖 5 所示為本研究提出之尋優支持向量機制分類模式以線性核函數的支持向量機的模擬退火搜尋演化歷程圖。由圖 5 可知，大約在 175 疊代的演化中便可得本實驗的最佳解；在表 1 指出在測試資料集中其分類錯誤率設為 0.1934。而圖 6 所示為本研究提出之尋優支持向量機制分類模式以半徑基底核函數之支持向量機的模擬退火搜尋演化歷程圖。由圖 6 可知大約在 250 疊代的演化中便可得本實驗的最佳解，而其在測試資料集中其分類錯誤率在表 2 可知為 0.2215。

表 1、SA-SVM 實驗最佳參數數據理(linear)

Support Vector Classification:
Constructing ...
Optimising ...
Execution time: 55.5 seconds
Status : OPTIMAL_SOLUTION
w0 ^2 : 1575274357731.632600
Margin : 0.000002
Sum alpha : 1575312485034.961700
Support Vectors : 214 (95.1%)
error_rate = 0.1934
val = 20.0440
[Use of Simulated annealing to find optimum]
>>>Best PI= 20.044048
x(1)= 5.361037、x(2)=224.926696
Training Dataset:255、Testing Dataset:423

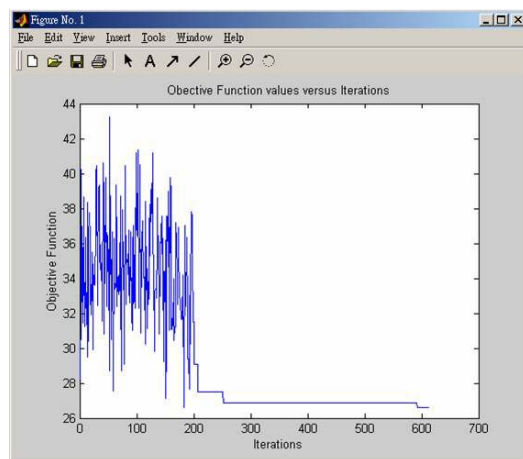


圖 6、SA-SVM 之模擬退火搜尋過程(rbf)

在醫療診斷糖尿病標準資料集的實驗中，本研究提出的尋優支持向量機制分類模式中在不同的核函數的支持向量機中其分類的錯誤率在線性模式的核函數為 0.1934 而利用模擬退火演算法中所搜尋的訓練資料集中有 95.1% (214/225) 為支持向量。另外以半徑

基底的核函數其分類的錯誤率為 0.2215 而利用模擬退火演算法中所搜尋的訓練資料集中有 81.8% (233/285) 為支持向量。其分類效能均比表 3 所列出的各種演算法為佳，可見本研究提出之尋優支持向量機制分類模式可改善支持向量機在大量的資料集中所浪費的資源並節省其計算的時間。

表 2、SA-SVM 實驗最佳參數數據理(rbf)

```
Support Vector Classification:
Constructing ...
Optimising ...
Execution time: 41.7 seconds
Status : OPTIMAL_SOLUTION
|w0|^2      : 117403.909687
Margin      : 0.005837
Sum alpha  : 117403.909626
Support Vectors : 233 (81.8%)
error_rate = 0.2215
val = 22.7821
[Use of Simulated annealing to find optimum]
>>>Best PI= 22.782115
x(1)= 24.243752、x(2)=284.635916
```

表 3、diabetes-實驗數據分類效能比較表

Algorithm	Train Error Rate	Test Error Rate	Train Time	Test Times
IGNF	0.172	0.208	9468	1
LogDisc	0.219	0.223	31	7
Dipol92	0.220	0.224	36	1
Discrim	0.220	0.225	27	7
Smart	0.177	0.232	3762	?
Radial	0.218	0.243	5	0
Itrule	0.223	0.245	31	2
BackProp	0.198	0.248	7171	0
Cal5	0.232	0.250	237	0
Cart	0.227	0.255	30	1
Castle	0.260	0.258	35	5
QuaDisc	0.237	0.262	24	7

<http://www.liacc.up.pt/ML/statlog/datasets/diabetes/diabetes.use.html>

六、結論與建議

本研究利用模擬退火演算法配合支持向量機，有效地藉由分類特徵，克服人工智慧技術在範例學習中設計的瓶頸。從實例驗證的結果顯示：本研究提出的「尋優支持向量機」，在分類應用上的效能，明顯優於之前學者所提出之其他演算法，亦可顯示本研究模式為一有效的混合型計算式智慧技術。

在後續的研究中，尚可運用其它有效的啟發式搜尋法，諸如遺傳演算法、禁忌搜尋法、增援演算法 (Reinforcement algorithm) ...等

等，建構更佳的分類應用模式，以擴展人工智慧運用的範疇。

七、參考文獻

- [1] 徐君豪，「全域最佳化之模擬退火法」，淡江大學機械工程研究所碩士論文，1998。
- [2] 蘇志傑，「模擬退火法之應用與改進」，碩士論文，國立中興大學機械工程學系，1999。
- [3] A. I. Abdelmola, and S. M. Taboum, "Productivity model for the cell formation problem: a simulated annealing algorithm," *Computers & Industrial Engineering*, 37, 327-330, (1999).
- [4] M. M. Ali, A. Torn, and S. Viitanen, "A direct search variant of the simulated annealing algorithm for optimization involving continuous variables," *Computers & Operation Research* 29, pp. 87-102, 2002.
- [5] K. P. Bennett and C. Campbell, "Support vector machines: hype or hallelujah?" *ACM Special Interest Group on Knowledge Discovery and Data Mining (SIGKDD) Expolarations*, Vol. 2, No. 2, pp. 1-13, 2000.
- [6] V. Chapelle, O. Vapnik, Bousquet, and S. Mukherjee, "Choosing multiple parameters for support vector machines," *Machine Learning*, Vol. 46, No. 1, pp. 131-160, 2002
- [7] A. Corana, M. Marchesi, C. Martini, and S. Ridella, "Minimizing multimodal functions of continuous variables with the simulated annealing algorithm," *ACM Trans. On Mathematical Software*, 13, 1987, pp.262-280.
- [8] C. Gold and P. Sollich, "Model selection for support vector machine classification," *Advances in Neural Information Processing Systems*, Vol. 15, 2002.
- [9] T. Joachims, "Making large-scale support vector machine learning practical advances in kernel methods—support vector learning", pp. 169-184, Cambridge, MA, 1999, MIT Press.
- [10] S. S. Keerthi and E. G. Gilbert. "Convergence of a generalized SMO algorithm for SVM classifier design", *Machine Learning* 46, 351-360, 2002.
- [11] J. C. Platt, "Probabilistic Outputs for Support Vector Machines for Pattern Recognition," U. Fayyad, Editor, 1999, *Kluwer Academic Publishers: Boston*