植基於支援向量機之資訊分享決策支援技術

劉江龍
國防大學理工學院
副教授
e-mail：
clliu@ndu.edu.tw

曾子軒
國防大學理工學院
研究生
e-mail：
c2981306@ndu.edu.tw

韓康年
國防大學理工學院
研究生
e-mail：
c2981204@ndu.edu.tw

婁德權
國防大學理工學院
教授
e-mail：
dclou@ccit.edu.tw

摘要

資訊分享可以提供組織早期獲得資安事件的預警，對資安事件的防範具有極大的幫助。因此，發展資訊分享平台已成為世界先進國家在防範資安事件方面的重要目標。本研究提出一有效的機制來支援決策者進行資訊分享之決策，稱為資訊分享決策支援技術，本研究並實際開發一分享決策支援系統，以驗證本技術的可行性。

實驗結果證明，本研究提出之資訊分享決策支援技術可以有效達成本技術設定的兩項功能需求：(1) 本技術可以在合理的時間內提供決策支援；(2) 本技術對於異常決策結果可以提供有效之方法進行決策模型之修正。而系統實作之結果證明本研究所提出之資訊分享決策支援技術之可行性。

關鍵詞：資安事件、資訊分享、決策支援技術，決策支援系統。

Abstract

Information sharing can provide organizations with early alarms for information security incidents. This functionality can help organizations to prevent the occurrence of information security incidents. Therefore, developing information sharing platform has become an important target of information security incident prevention for the developed countries in the world. This study proposes an effective information sharing decision support technique (called ISDS technique) to support the decision maker to make decisions for information sharing. This study also develops an information sharing decision support system to show the effectiveness of the proposed technique.

Experimental results show that the proposed ISDS technique can meet the following requirements: (1) The proposed technique can provide the decision support in reasonable time; (2) For the abnormal decisions, the proposed technique can provide effective method to correct the decision model. Moreover, the implementation of the proposed system also proves the feasibility of the proposed technique..

Keywords: **Information security incidents, information sharing, decision support technique, decision support system**

1. 前言

網路的進步加速了知識的傳播，但也同時加速病毒及惡意程式的漫延。因此，若能在資安事件發生之初，透過有效的分享機制將資安事件進行分享，則其他仍未被資安事件影響的組織即可早期獲得預警，從而整合各種資安事件情報，並進行分析、研討、再準備最佳的應對方式，如此對杜絕資安事件的漫延將有很大的助益。

資訊分享及分析中心(ISACs, Information Sharing and Analyzing Centers)起源於 1998 年美國 63 號總統決策令（PDD-63）[1]，PDD-63 中要求各政府部門需與相關之民間產業間建立合作關係並分享資訊，以利共同協助保護美國關鍵基礎建設。2003 年國土安全總統 7 號指令（HSPD-7）[2]更修正了 PDD-63，重申各政府部門應推動與相關民間產業建立資訊分享及分析中心(ISACs)，主要任務包括確認、辨識及整合關鍵基礎建設之保護措施，並推動弱點、威脅及事件的資訊分享。自 1999 年至今，美國已主導成立約 14 多個行業類別之 ISAC，其推動的成效也已引起歐盟、英國等國設置各國 ISAC。

有鑒於國內資通安全發展現況及未來趨勢，我國「國家資通安全會報」自 94 年開始推動第二期「建立我國通資訊基礎建設安全機

制計畫(94 年至 97 年)」[3]，其中訂定了 10 項發展目標，並且在其第 2 項明訂「建置政府及重要基礎建設之資訊分享及分析中心，提昇國家競爭力」。為防範各式資安事件，行政院下轄國防、外交、交通跟經濟四大部會，將優先建置資訊、通訊的基礎建設安全機制。

由於不同資安事件對不同性質的組織來說，其機敏性有所不同，加上目前沒有好的機制被提出來進行分享資訊的過濾，因此，大部分組織在資訊安全的考慮下，對是否進行內部資安事件之分享有所顧慮。圖 1 為在具有資安顧慮下可能之 ISAC 資訊分享概念，其中事件資料庫的來源可能包括組織內部提供的分享及外部主動情蒐、國際交流、或產業合作所獲得的資料，這些資料經由組織內部分享決策機制決策後，再透過分享平台機制將值得分享之資料分享給外部 ISAC 單位，包括 N-ISAC、行政院 ISAC、國防產業 ISAC、及民間產業 ISAC 等。

根據上述之分享概念，最後分享給外部的資訊是已管控處理後的資安事件，而組織內部分享決策機制即在對來自不同來源之事件進行分享之決策。然而在內部產生之事件中不乏機敏之事件，為了同時達到資料管控及資訊交流的目的，在資訊分享之前必須先行對資訊進行分享決策，以期達到安全資訊分享的目的。本研究之目的即在提出一有效之機制以支援決策者進行資訊分享之決策。

依據本研究之規劃，資訊分享決策機制共分為兩個部分進行，一為事件管理機制，另一為分享決策支援機制。事件管理機制的目的在針對匯入事件資料庫之資安事件產生特徵屬性(Characteristic Attributes)，並對特徵屬性進行量化(Ranking)。量化的目的是以數字來對特徵屬性賦予等級，以方便事件的區別及整理。而分享決策機制即是將經由事件管理機制整理好的資訊進行後續處理，以決定事件的分享方式。本論文著重於分享決策機制的研究。

對於決策者而言，其分享決策不外分為兩種，一種為外部分享，一種為內部分享。從事件決策機制角度來看，所謂外部分享，是經由分享決策機制判定為可分享的事件；而所謂內部分享，則是則是經由分享決策機制判定為不可分享的事件。為了有效支援決策者進行決策，本研究目的在提出一分享決策支援技術，並設定本決策支援技術必須達到下列兩項功能需求：

(1) 決策支援技術必須在合理的時間內提供決策支援。所謂合理的時間，表示決策者可容忍的等待時間。
(2) 對於異常決策結果，必須提供有效之方法來進行修正。所謂異常決策，表示經由決策支援技術判定為可分享事件，但決策者最後判定為不可分享事件，或者經由決策支援技術判定為不可分享事件，而決策者最後判定為可分享事件。

目前有許多種分類技術可用來進行上述之決策，較為常用的為類神經網路(Neural Network) [4, 5, 6]、支援向量機(SVM, Support Vector Machine) [7, 8, 9]、及模糊理論(Fuzzy Theory)[10, 11]等。其中支援向量機對於二分法的決策具有良好的穩定度及成熟度，較符合本研究之需求，因此本研究所提分享決策支援技術採用支援向量機來進行分享決策。

為完整介紹本研究，本論文其餘各節安排如下：第 2 節介紹支援向量機的運作原理；第 3 節介紹本研究所提之資訊分享決策支援技術；第 4 節說明與本技術有關的各項實驗結果；第 5 節為本論文的結論。

2. 支援向量機簡介

支援向量機是 1995 年由英國倫敦 Royal Holloway 大學電腦科學系教授 V. Vapnik 及 AT&T 實驗室研究團隊所提出的方法[12]。其理論基礎是利用統計學習理論中結構風險最小誤差法及分類邊界最大化原則，在類別間利用分類超平面當作區隔標準來區分兩種或多種不同類別的集合，以及透過訓練功能達到學習任務，進而得到最佳的推廣性能。支援向量機早期是用來發展二分法的分類應用，後來逐漸發展成多分類的應用，目前被廣泛運用於各種領域，例如圖像分類、手寫字體識別、人臉識別、影像識別、資料探勘、三維目標識別等。

簡單來說，支援向量機需要找出一個超平面(Hyperplane)，能夠將兩個不同的集合分開。因為實際資料可能是屬於高維度的資料，因此超平面意指在高維中的平面。以二維資料為例(如圖 2 所示)，黑點及白點分別代表不同資料集合，X 軸及 Y 軸則代表兩種不同屬性。我們希望在圖中找出一條線，其能夠將黑點和白點分開(分類)，而且這條線距離這兩個集合的邊界(margin)越大越好，這樣才能夠明確的分辨這個點是屬於那個集合，否則在計算上容易產生誤差。

為了讓接下來的支援向量機說明更順利，以下定義了幾個符號：

(1) x_i ：為向量(Vector)，表示單一筆資料的個別屬性(Attribute)，$x_i \in R^N$，i=1,2,…l。

(2) y_i：表示x_i類別，$y_i \in \{ \pm 1 \}$，i=1,2,…l。

(3) w：向量權重值。

(4) f：決策函數(Decision Function)，$f(x_i) = sign(w \cdot x_i + b)$，$f: R^N \rightarrow \{ \pm 1 \}$，決策函數可以決定$x_i$是屬於那一分類(+1或−1)。

SVM主要的目的是使用訓練資料找出一個超平面來分出它們所屬的類別，亦即求出 $f: wx + b = 0$ 這個超平面所對應的權重w與係數b。當新的資料x_j進來時，我們就能使用決策函數 $f(x_i) = w \cdot x_i + b$ 來對它進行分類。如果 $f(x_j) > 0$，則x_j屬於類別(+1)；如果 $f(x_j) < 0$，則x_j屬於類別(−1)。這就是支援向量機運作原理。

支援向量機可概略區分為線性(Linear)及非線性(Non-linear)兩大類。線性支援向量機[13]是支援向量機中較簡易的種類，對於線性的資料集具有良好的分類效果，原理也較容易了解。對圖2而言，$w \cdot x_i - b$代表分類超平面，因此，圖中的右上半部白點區塊被分類到(+1)，以數學表示如下式：

$$w \cdot x_i - b \geq +1 \quad \forall y_i = +1 。 \tag{1}$$

而圖中左下半部黑點區域則被分類側(−1)，以數學表示如下式：

$$w \cdot x_i - b \leq -1 \quad \forall y_i = -1 。 \tag{2}$$

結合(1)及(2)式，可得

$$y_i(w \cdot x_i - b) - 1 \geq 0 \quad \forall i 。 \tag{3}$$

由上式及圖 2 知，分類超平面與兩邊界距離為 $2/\| w \|$，此時兩邊界稱為支援超平面(Support Hyperplane)，兩支援超平面則越接近各分類區塊越好。被分隔的分類區塊之間的距離稱之為 Margin(如圖 3 所示)。Margin 在不重疊到資料點的前提下越大越好，如此分類超平面對於預測新資料會越準確。如在圖 3 中，當分別在左圖及右圖中新加入預測點，並保持其與各分類點位置不變，我們可以發現，其將因不同 Margin 條件而有不同的分類結果。在左圖中，理應分類到白點區域的新預測點，卻因 Margin 過小的超平面而產生分類誤差，導致分類至黑點區域。

要求Margin距離最大值時，可在公式(3)條件下使用Lagrange Multiplier Method求 $\| w \|^2$ 最小值，此時在支援超平面上的x_i稱為支援向量。經轉換(3)式後可得

$$L_p(w,b,a) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{l} a_i[y_i(w \cdot x_i - b) - 1] 。 \tag{4}$$

對w及b分別進行偏微分後，可得

$$\frac{\partial}{\partial w} L_p \Rightarrow w = \sum_i a_i y_i x_i ， \tag{5}$$

$$\frac{\partial}{\partial b} L_p \Rightarrow \sum_i a_i y_i = 0 。 \tag{6}$$

將(5)及(6)式代回(4)式後，可得一新方程式

$$L_D = \sum_i a_i - \frac{1}{2}\sum_{i,j} a_i a_j y_i y_j x_i \cdot x_j ， \tag{7}$$

Lagrange Multiplier Method：$a_i \geq 0$，

Complementary slakness：

$a_i[y_i(w \cdot x_i - b) - 1] = 0$ 。

上述條件又稱為KKT條件(Karush Kuhn Tucker conditions)。KKT條件是在數學中一個非線性規劃問題，能有最佳化解法的一個充分必要條件，這是一個廣義的Lagrange Multiplier Method成果。在實施訓練時，訓練資料中有些點會滿足KKT條件，這些點是位於支援超平面上，又稱為支援向量，上述式中可確定 $\| w \|^2$ 最小值，也可確定分類超平面位置，此時超平面可用來判斷新預測的點是屬於那個集合，完成分類。

線性支援向量機是以線性函數區分兩類不同資料，在處理分類與預測問題時，隨著未處理資料的複雜度及資料量的增加，線性函數分類並不一定適用於所有的資料。因此若遇到非線性的問題，就可以使用非線性支援向量機來解決。

非線性支援向量機的作法是先將原始資料透過映射函數Φ轉換，將原始的資料空間映射到另一個維度的特徵空間(Feature Space)(如圖4所示)，使資料較容易以線性函數來進行區分，並且在特徵空間中尋求線性最佳分類超平面，可以得到更好的正確率。而在做這個非線性轉換之前必須事先設定核心函數（Kernel）。

資料映射到特徵空間需透過Φ映射函數轉換，亦即$x_i x_j \Rightarrow \Phi(x_i) \cdot \Phi(x_j)$，而核心函數可定義為 $K(x_i,x_j) = \Phi(x_i) \cdot \Phi(x_j)$，在之前的數學推導中$x_i x_j$都是湊在一起呈現內積形式，因為Φ映射函數可能為很複雜繁瑣的函數，但是內積形式卻可以很簡單，所以不需針對x_i或x_j及$\Phi(x_i)$或$\Phi(x_j)$單獨求值，只要針對$\Phi(x_i) \cdot \Phi(x_j)$求值即可。

核心函數有許多種類，本節僅列出常用四種形式：

(1) 線性(Linear)：

$$K(x_i, x_j) = (\Phi(x_i) \cdot \Phi(x_j)) 。 \tag{8}$$

(2) 多項式(Polynomial)：

$$K(x_i, x_j) = (1 + x_i \cdot x_j)^p 。 \tag{9}$$

(3) 放射基底(RBF, Radial Basis Function)：

$$K(x_i, x_j) = \exp(\frac{-\|x_i - x_j\|^2}{2\sigma^2}) \text{ 。} \quad (10)$$

(4) S 型(Sigmoid)：

$$K(x_i, x_j) = \tanh(kx_i \cdot x_j + \Theta) \text{ ，} \quad (11)$$

其中 k 為增益(Gain)，Θ則為門鑑值(Threshold)。

在應用上，上述四種核心函數中以 RBF 平均表現準確度較好，因此，本研究採用 RBF 核心函數來進行實驗及系統設計。

3. 基於支援向量機之資訊分享決策支援技術

3.1 系統架構

本節說明本研究所提出的資訊分享決策支援技術，簡稱本分享決策支援技術。由於本分享決策支援技術的目的在發展一資訊分享決策支援系統，因此，在未來各節中，本論文稱利用本分享決策支援技術所發展之系統為「資訊分享決策支援系統」，簡稱為本系統。

圖 5 為本分享決策支援系統架構圖。如第 1 節所述，來自外部或內部的事件資料會先匯入事件資料庫，並經過事件管理系統處理後，產生關鍵屬性，連同原來事件屬性，一併匯入中介資料庫。因此，在中介資料庫中的事件資料即為本系統之輸入。本系統則利用一事先建好的決策模型(Decision Model)，對中介資料庫中新產生的事件進行決策，並將決策結果利用一決策支援介面，提供給決策者作最後的決策。若決策者認為事件可以分享，則系統會根據各事件屬性產生分享屬性，並同時複製到外部分享資料庫及內部分享資料庫中；若決策者認為事件不可以分享，則系統會根據各事件屬性產生分享屬性，並將其複製到內部分享資料庫中。

本系統採用之基礎分享支援技術之系統流程如圖 6 所示。如前所述，在中介資料庫中，每筆事件均有多個屬性，其是由事件管理系統從原本完整的資安事件敘述或屬性產生為關鍵屬性，並予量化後，併同原來事件屬性儲存於中介資料庫中。接著可由具資安相關工作經驗的管理者選出一些可確定分類的代表性事件，並透過 SVM 分類器訓練後，產生決策模型。當一筆新的資安事件存到中介資料庫中，系統則根據訓練好的決策模型判斷該事件是否適合分享，並將判斷結果提供決策者作為決策之依據。最後資安決策者以人工方式決定分類判斷是否恰當，如果決策者同意系統對於該筆事件之判斷，該筆事件即會轉存入分享資料庫實施分享，如果分類結果不如決策者預期，此筆事件稱為異常決策事件(Abnormal Decision Incident)，代表系統決策模型仍有調整空間，需要修正決策模型，該事件則回饋給系統，系統則將原始模型所採用之訓練事件加上此筆異常事件，重新進行分類器訓練，產生新的決策模型。而系統將根據新的訓練模型進行下一筆新事件的決策支援。

根據 SVM 原理，當訓練事件樣本逐漸增大時，加入單一筆異常決策事件所產生的新決策模無法更改具有類似屬性的決策，因此，在實務上，本研究則利用事件複製技術進行新決策模型的產生。有關此部分的討論，則留到第 4 節說明。

3.2 系統使用資料庫

本系統使用數個資料庫，包括中介資料庫、外部分享資料庫、內部分享資料庫、訓練與異常事件資料庫、事件模型資料庫、以及使用者帳號密碼資料庫等。所有資料依其屬性分別儲存於不同資料庫中，以互動式網頁介面連結使用。

中介資料庫中的事件屬性規劃與各屬性分級將直接關係到支援向量機的判斷準確度。所謂事件屬性即是一筆事件中可以歸納統計整理的部分，對不同的事件來說，其所能統計出的屬性項目也不盡相同，隨著事件樣本的增加，屬性欄位的訂立也會愈加明確。事件的關鍵屬性即是本系統支援向量機用以訓練決策模型的重要依據，一筆事件中可篩選出的關鍵屬性越多及等級評定越準確，分類準確率就越高。但是關鍵屬性訂立過多，不但會拖累系統速度，且過多的關鍵屬性對於一些事件而言，反而會形成過多的空值(Null)，造成準確率的不確定，所以資料庫中的屬性欄位的訂定需詳加考慮。經過前端事件管理系統測試後，初期訂出五個事件關鍵屬性作為 SVM 訓練器的輸入項目，分別是資產等級、機密等級、威脅等級、系統修補程度、弱點等級等五個屬性，如表 1 所示。

資產等級依據事件本身或事件影響目標的資產價值於予區分為：1(最低) 、2(低)、3(中)、4(高)、5(最高)等五個等級，分別對應到 SVM 的屬性數值 1~5；機密等級依據事件的機密屬性區分為：1(普通)、2(密)、3(機密)、

4(極機密)等四個等級，分別對應到 SVM 的屬性數值 1~4；威脅等級依據事件本身或對事件影響的目標威脅區分為：1(最低)、2(低)、3(中)、4(高)、5(最高)等五個等級，分別對應到 SVM 的屬性數值 1~5；系統修補程度則區分：1(已修補)、2(未修補)等二個等級，分別對應到 SVM 的屬性數值 1~2;弱點等級依造成事件的弱點程度區分為：1(無)、2(低)、3(中)、4(高)等四個等級，分別對應到 SVM 的屬性數值 1~4。所以中介資料庫欄位雖多，但真正輸入 SVM 的每筆事件的屬性值只有六個向量(Vector)（五個關鍵屬性加上一分類類別）。

內部及外部分享資料庫中之資料為中介資料庫之精簡版，如表 2 所示，僅儲存需要分享給使用者的資料欄位，其中事件摘要說明欄位是由其他欄位組合而成的簡易的事件敘述，系統設定事件的最後決策若為可分享，則精簡版的事件資料會同時存入內部或外部兩資料庫；如果事件的最後決策為不分享，則精簡的事件資料則只會存入內部分享資料庫，不會存入外部分享資料庫。也就是說只要事件經過處理，不管分享與否均會出現在內部資料庫。而系統真正要分享給部外資訊是儲存於外部資料庫中的資料。

4. 實驗結果

為證明本研究提出之分享決策技術的有效性，本研究根據本分享決策支援技術實際開發軟體進行實驗與測試。本研究設定最後系統的執行環境是以主從架構 (Client-Server Architecture) 執行。因此，在系統開發初期採用單機程式開發與測試，亦即將資料庫與程式開發平台安排在同一部電腦中，在系統開發成熟後，再與前端事件管理系統進行整合測試。

為了避免因前端事件管理系統的誤差而影響實驗結果，本研究採取人工方式建造事件資料庫，模擬事件的蒐集來源大部分為行政院國家資通安全會報及資通安全資訊網[14]等發佈的資安事件，自 2006 至 2007，共蒐集約 500 筆事件資料，其中包含資安事件、病毒警告、資安技術通報、資訊犯罪報導等各式資料。事件的關鍵屬性的級數均由具資安管理經驗者手動賦予，並經嚴格檢測與交叉比對，以確保事件資料量化等級的準確性。同時由具資安管理經驗者對各事件先行進行分享決策，此決策與關鍵屬性的分級資料則透過知名的 LibSVM 程式[15]進行決策模型的訓練，同時選擇 RBF 作為基底函數。為了實驗需要，本研究共完成 50 筆、100 筆、250 筆、及 500 筆樣本事件所訓練的決策模型，分別簡稱為 50 筆決策模型、100 筆決策模型、250 筆決策模型、及 500 筆決策模型等。

為符合本研究設定之功能需求，本實驗區分為四個部分：第一部分為決策模型準確度測試，主要是測試本系統所製作決策模型判斷的正確率；第二部分為系統反應測試，主要是測試本系統是否可以在合理的時間內提供決策支援；第三部分為決策模型對未知事件判斷準確度測試，主要是測試不同數目的決策模型對未知事件判斷的準確度；第四部分為決策模型自動修正功能測試，主要是測試本系統設計的自我決策模型修正的能力。以下各小節說明上述個別的實驗結果。

4.1 決策模型準確度測試

為測試決策模型的準確度，本實驗利用 50 筆、100 筆、250 筆、及 500 筆代表事件分別產生四種決策模型，並對各模型之原始訓練事件進行準確率實驗，實驗結果如表 3 所示。由表 3 中可得知，50 筆及 100 筆決策模型對於原訓練事件之判斷的準確率雖高，但皆有一筆判斷錯誤(準確率分別為 98%及 99%)，亦即該錯誤判斷事件原始設定為不可分享(分類數值為 –1)，但由 50 筆及 100 筆決策模型判斷後卻為可分享(分類數值為 1)。究其原因，是因為事件的關鍵屬性共有 800 種組合，而此兩部分決策模型是由 50 筆及 100 筆樣本事件所訓練而成，顯然不足，造成分類超平面不夠準確所致。而 250 筆及 500 筆決策模型則有較佳之表現，準確率達 100%。

4.2 系統反應測試

為測試系統的反應時間，本實驗分別利用上述 50 筆、100 筆、250 筆、及 500 筆決策模型對相同 10 筆未知事件進行決策，平均反應時間如表 4 所示。由表 4 中可明顯發現，不同決策模型對相同事件的平均反應時間均在 0.2 秒左右，可稱為是即時的反應。因此，本系統滿足在系統設定的可在合理時間內提供決策支援的需求。

4.3 未知事件判斷準確度測試

為測試決策模型對未知事件判斷的準確度，本實驗延續第一部分實驗，分別利用 50 筆、100 筆、250 筆、及 500 筆決策模型對未知事件實施判斷，實驗結果如表 5 所示。由表

5 中可知，50 筆決策模型對於 50 筆、100 筆、250 筆、及 500 筆測試事件的準確率表現均未能達到 100%，而 100 筆決策模型對測試事件準確率的表現則較 50 筆決策模型高，但是仍非最好表現，顯示用以訓練兩個模型的事件筆數仍屬不足，而 250 筆及 500 筆決策模型則有著較佳之表現，此部分結果驗證了第一部分實驗的概念。

4.4 決策模型自動修正功能測試

本實驗在測試決策模型自動修正所需事件數，也就測試異常決策事件與模型自動修正的關係，這部份之測試結果將決定系統設定自動修正之條件參數。每一事件會因時間、地點、內容不同而具有唯一性，但是不同事件也許會有相同關鍵屬性值，本實驗即在測試需要多少重複的異常決策事件才可以正確改變決策模型的判斷。為完成此部分實驗的目的，本實驗直接在訓練資料庫中逐步加入相同異常決策事件，藉以修正決策模型，再以此決策模型針對同數值事件加以判斷。為了不失一般性，本實驗分別對一般異常事件(亦即位於分類超平面附近的異常事件)及極端異常事件(亦即位於邊緣區域之事件)進行測試。

在一般異常事件實驗部分，本實驗選定在第一部分實驗中位於分類超平面附近的異常決策事件 (關鍵屬性值為 24431)作為測試用事件，表 6 說明影響決策模型所需單一事件複製筆數之實驗結果。其中數值為 1 的部分是原始判斷該事件為可分享的結果，而數值為-1 的部分是我們希望修正後的模型的對於該事件判斷的結果（即不可分享）。由表 6 中可以看出，當測試異常決策事件筆數為 1 筆時，無法完全對於各決策模型產生影響，但累計加入訓練資料庫 3 筆以上異常決策事件時，判斷結果就可達到我們所期望的修正值。

我們除了希望修正後的模型能有不同的判斷結果外，我們也不希望修正後的決策模型影響到對原來事件判斷的準確率。為測試加入異常決策事件的模型是否會影響到對原始訓練事件判斷的結果，本研究同時使用新決策模型對原模型使用訓練事件進行判斷，實驗結果如表 7 所示。從表 7 中可看出，除了 100 筆決策模型有 1 筆判斷錯誤外(正確率 99%)，其餘決策模的判斷正確率均為 100%。

在極端異常決策事件實驗部分，本實驗以關鍵屬性為 11111 極端點作為測試事件，其實驗過程均與上述一般異常決策事件實驗部分相同，其結果如表 8 及 9 所示。由表 8 可看出，對極端異常決策事件來說，如果加入異常決策事件的點數不多，則新決策模型無法進行決策之變更，可是當加入異常決策事件增多時，將會改變決策模型之決策。例如加入 5 筆極端異常決策事件時，只有 50 筆及 100 筆決策模型的判斷結果會被影響；但當加入 20 筆異常決策事件時，則含 250 筆以下之決策模型的決策結果均會被影響。另外從表 9 中可看出，在極端狀況下，若以新建立的決策模型回測原始訓練事件的判斷準確率，則 100 筆以下的決策模型則顯出樣本不足的現象，因此容易被新加入的極端異常決策事件所影響，其穩定度沒有 250 筆決策模型來得穩定

5. 結論

建立資訊分享及分析中心為目前行政院重要推動的政策。根據 ISAC 資訊分享概念，最後分享給外部的資訊是已管控處理後的資安事件，而組織內部分享決策機制即在對來自不同來源之事件進行分享之決策。為了同時達到資料管控及資訊交流的目的，在資訊分享之前必須先行對事件進行分享決策，以期達到安全的資訊分享目的。

本研究提出一有效之機制來達到決策支援之目的。本研究主要是利用支援向量機對歷史決策的資安事件進行訓練，產生決策模型，再依據決策模型對未知事件進行決策，提供決策者作為新資安事件分享與否決策之依據。為有效建構「資訊分享決策支援系統」，本研究提出一基於 SVM 的分享決策支援技術。

實驗結果證明，本研究提出之「資訊分享決策支援系統」可以有效達成系統所設定之功能目標，包括：

(1) 系統必須在合理的時間內提供決策支援。經實驗結果，本系統可以在平均 0.2 秒內提供新事件的決策支援。
(2) 對於異常決策結果，系統必須提供有效之方法進行修正。本研究提出模型自動修正機制，利用異常決策事件回饋方式產生機器學習的能力，隨著使用者針對事件判斷的標準轉變，系統也會逐步修正決策模型以符合決策者理念，進而提升決策的準確率。

致謝

本研究為中華民國行政院國家科學委員會專題研究計畫部分成果，計畫編號：NSC

96-3114-P-606-002-Y。

參考文獻

[1] Critical Infrastructure Protection (PDD 63), http://fas.org/irp/offdocs/pdd/pdd-63.htm.

[2] December 17, 2003 Homeland Security Presidential Directive/Hspd-7, http://www.whitehouse.gov/news/releases/2003/12/20031217-5.html.

[3] 建立我國通資訊基礎建設安全機制計畫，http://www.pthg.gov.tw/CmsFile/200742694854234.pdf。

[4] Rowley, H. A., Baluja, S., and Kanade, T., "Neural Network-Based Face Detection," ***IEEE Transactions on Pattern Analysis and Machine Intelligence***, Vol. 20, No.1, pp. 203–208, Jan. 1998.

[5] Samanwoy, G.-D., Adeli, H., and Dadmehr, N., "Principal Component Analysis-Enhanced Cosine Radial Basis Function Neural Network for Robust Epilepsy and Seizure Detection," ***IEEE Transactions on Biomedical Engineering,*** Vol. 55, No. 2, pp. 512–518, Feb. 2008.

[6] Visa, A. and Iivarinen, J., "Evolution and Evaluation of a Trainable Cloud Classifier," ***IEEE Transactions on Geoscience and Remote Sensing***, Vol. 35, No. 5, pp. 1307–1315, Sept. 1997.

[7] Archibald, R. and Fann, G., "Feature Selection and Classification of Hyperspectral Image with Support Vector Machines," ***IEEE Geoscience and Remote Seneing Gletters***, Vol. 4, No. 4, pp. 674–677, Oct. 2007.

[8] Kamruzzaman, J. and Begg, R. K., "Support Vector Machines and Other Pattern Recognition Approaches to the Diagnosis of Cerebral Palsy Gait," ***IEEE Trasactions on Biomedical Engineering***, Vol. 53, No. 12, pp. 2479–2490, Dec. 2006.

[9] He, J., Hu, H.-J., Harrison, R.,Tai, P. C., and Pan, Y. "Rule Generation for Protein Secondary Structure Prediction with Support Vector Machines and Decision Tree," ***IEEE Transactions on Nanobioscience***, Vol. 5, No. 1, pp. 46–53, Mar. 2006.

[10] Liang, Q. and Mendel, J. M., "MPEG VBR Video Traffic Modeling and Classification Using Fuzzy Technique," ***IEEE Transactions on Fuzzy System***, Vol. 9, No. 1, pp. 183–193, Feb. 2001.

[11] Ray, K. S. and Dinda, T. K., "Pattern Classification Using Fuzzy Relational Calculus," ***IEEE Transactions on System, Man, and Cybernetics-PART B***, Vol. 33, No. 1, pp. 1–16, Feb. 2003.

[12] Vapnik, V., ***Statistical Learning Theory***, Wiley, New York, 1998.

[13] 林宗勳，"Support Vector Machines 簡介"，http://www.cmlab.csie.ntu.edu.tw/~cyy/learning/tutorials/SVM2.pdf。

[14] 資通安全資訊網，http://ics.stpi.org.tw/index.html。

[15] LIBSVM--A Library for Support Vector Machines，http://www.csie.ntu.edu.tw/~cjlin/libsvm/ index.html。

表1 中介資料庫的關鍵屬性

屬性名稱	**說明**
資產等級	區分：1(最低) 2(低) 3(中)4(高)5(最高)
機密等級	區分：1(普通)2(密)3(機密)4(極機密)
威脅等級	區分：1(最低)2(低)3(中)4(高)5(最高)
系統修補程度	區分：1(已修補)2(未修補)
弱點等級	區分：1(無)2(低)3(中)4(高)

表2 內外分享資料庫使用之屬性

屬性名稱	**說明**
發佈編號	區分 A(內部)B(外部)C(情蒐)
發佈日期	發佈日期
事件日期	說明事件發生時間
事件類別	區分：1 (病毒)2(惡意程式)3(私接民網)4(其他)
事件地點	回報之事件發生地點不見得是回報單位地點。
事件摘要說明	摘要式事件說明
影響等級	事件本身所帶來的風險等級評估
影響系統	作業系統類型
採取措施	事件是否已完成修補，或是狀況是否已經復原
存取權限	區分:A(管理者)B(IT 人員)C(一般使用者)
完整事件敘述	完整事件敘述

表3 不同決策模型對個別訓練事件之決策準確率

	決策模型使用訓練事件筆數			
	50	100	250	500
準確率	98%	99%	100%	100%

表4 不同決策模型之平均反應時間

	決策模型使用訓練事件筆數			
	50	100	250	500
平均反應時間(秒)	0.199	0.198	0.199	0.201

表5 不同決策模型對未知事件測試結果

測試事件數	決策模型使用訓練事件數			
	50	100	250	500
50	98%	100%	100%	100%
100	96%	99%	100%	100%
250	96%	99.2%	100%	100%
500	97%	99.2%	100%	100%

表 6 一般狀況下，單一異常決策事件重複筆數影響決策模型情形

決策模型	單一異常事件重複筆數								
	1	2	3	4	5	10	20	30	40
50	1	1	-1	-1	-1	-1	-1	-1	-1
100	1	-1	-1	-1	-1	-1	-1	-1	-1
250	1	1	-1	-1	-1	-1	-1	-1	-1
500	1	-1	-1	-1	-1	-1	-1	-1	-1

表 7 一般狀況下，新決策模型對原訓練事件判斷準確率

決策模型	單一異常事件重複筆數								
	1	2	3	4	5	10	20	30	40
50	100%	100%	100%	100%	100%	100%	100%	100%	100%
100	99%	99%	99%	99%	99%	99%	99%	99%	99%
250	100%	100%	100%	100%	100%	100%	100%	100%	100%
500	100%	100%	100%	100%	100%	100%	100%	100%	100%

表 8 極端狀況下，單一異常事件重複筆數影響決策模型情形

決策模型	單一異常事件重複筆數								
	1	2	3	4	5	10	20	30	40
50	1	1	1	1	-1	-1	-1	-1	-1
100	1	1	1	1	-1	-1	-1	-1	-1
250	1	1	1	1	1	1	-1	-1	-1
500	1	1	1	1	1	1	1	1	1

表 9 極端狀況下，新決策模型對原訓練事件判斷準確率

決策模型	單一異常事件重複筆數								
	1	2	3	4	5	10	20	30	40
50	98%	98%	98%	98%	96%	98%	98%	98%	98%
100	98%	98%	98%	98%	98%	98%	98%	98%	98%
250	100%	100%	100%	100%	100%	100%	99.6%	99.6%	99.6%
500	100%	100%	100%	100%	100%	100%	100%	100%	100%

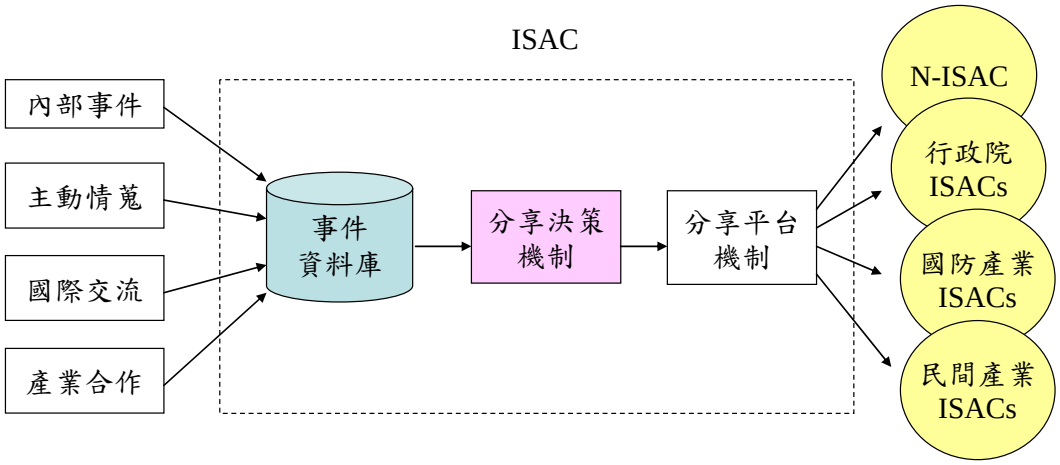

圖1 具資安顧慮下可能之ISAC分享概念

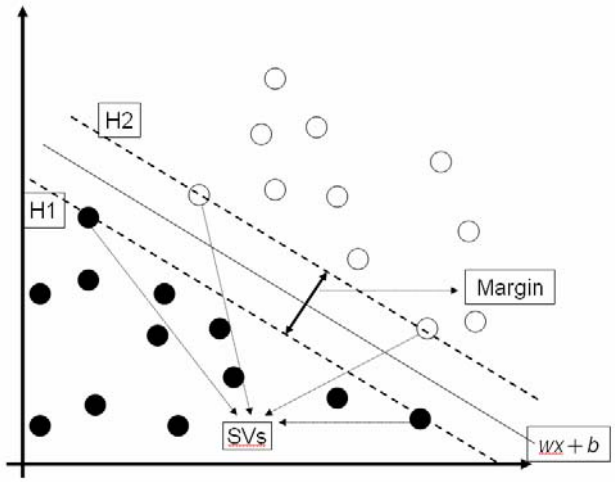

圖 2 SVM 分類示意圖

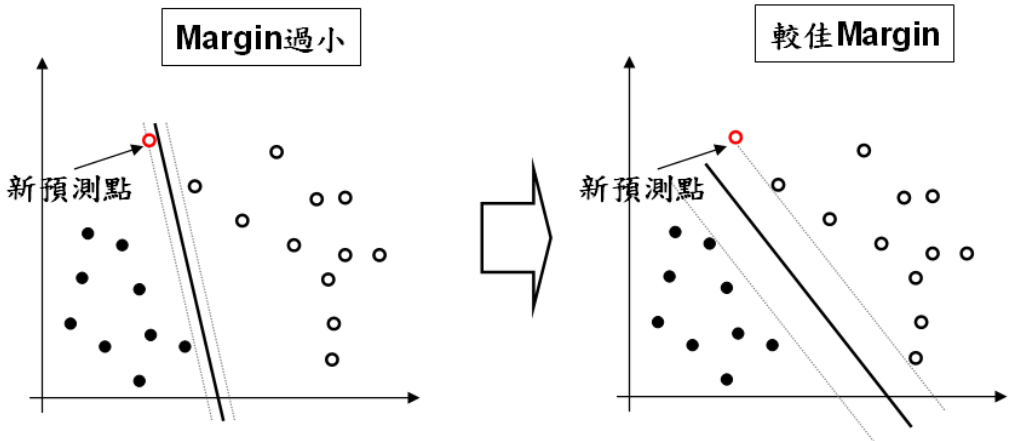

圖 3 不同 Margin 之分類差異圖

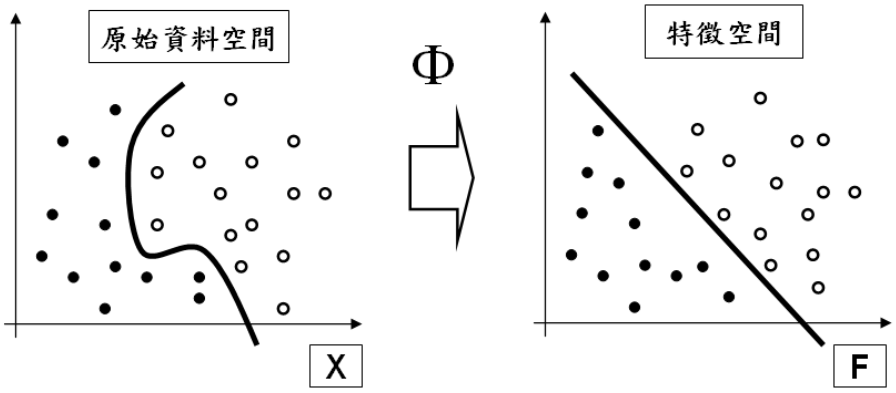

圖 4 二維原始資料轉換至特徵空間示意圖

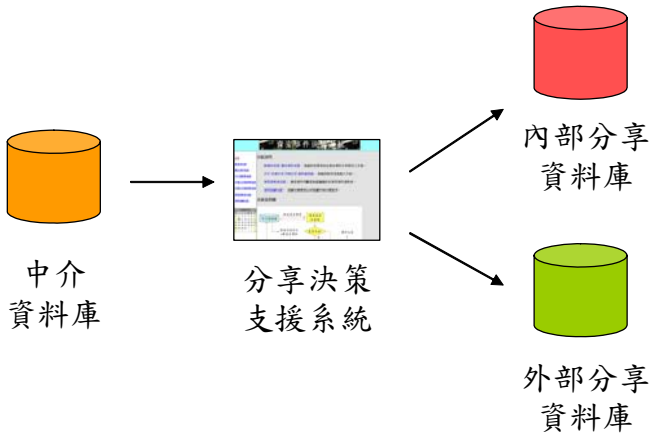

圖5 本分享系統架構圖

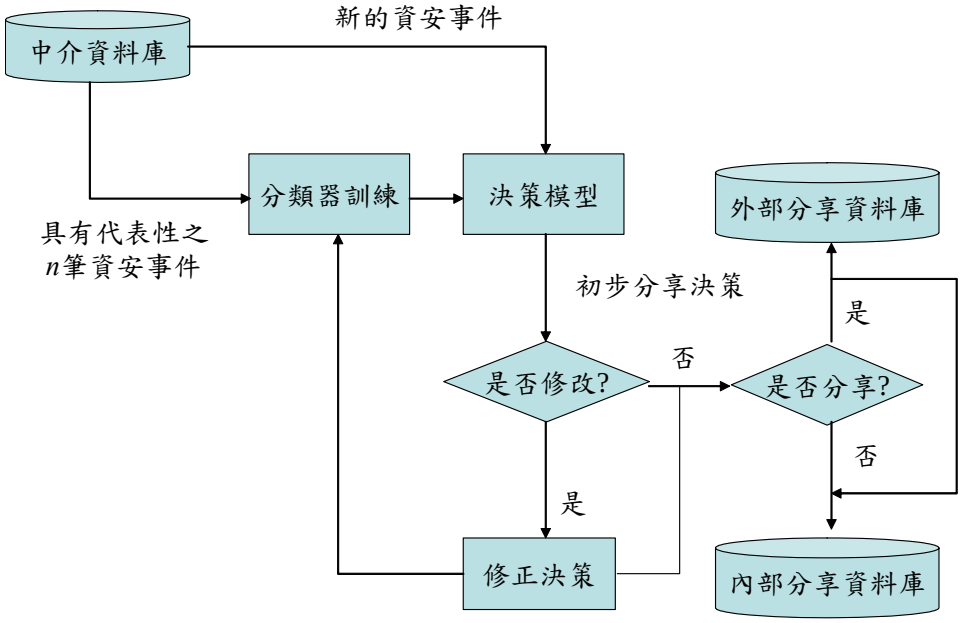

圖6 本分享決策技術流程圖