

以視訊內容為基礎之視訊資料分類 Content-based Classification for Video Data

劉敦仁
Duen-Ren Liu

林貞嫻
Chen-Hsien Lin

黃景彰
Jing-Jang Hwang

國立交通大學資訊管理研究所
Institute of Information Management
National Chiao Tung University

摘要

如何將視訊資料予以分門別類，藉以協助使用者瀏覽搜尋其有興趣的視訊資料，是重要的研究課題。本文探討視訊資料分類之自動化，以改進人工分類之缺點。本文提出視訊資料分類之方法，以視訊內容之關鍵字描述為基礎來進行分類。不同於文字文件之分類，視訊資料具有結構階層之特性，因此所提出的方法是依據視訊內容在視訊階層之權重值來分類。

關鍵字：視訊分類，視訊內容，資訊擷取

Abstract

It is a key issue to classify video data into categories, in order to help user browse and search video data. This paper shows how automatic classification of video data can be achieved, so as to amend the disadvantages of human classification. The paper proposes an approach to classify video data based on video content. Different from the classification of text documents, the classification of video data is according to the structural weight of video content.

Keywords : Video Classification, Video Content, Information Retrieval

1. 緒論

任何資料形態之文件檔案(包括文字、圖形、影像、視訊....等)的資料庫在建立時，都牽涉到一項重要的步驟：將文件檔案依據其討論的內容歸類到代表相關的主題或領域的類別，也就是分類(classification)的過程[2,5,6,7,11]。分類的目的在於讓對於某主題有興趣的使用者可以瀏覽擷取該類別的文件。而由於全球資訊網(World Wide Web)支援多媒體資料型態的技術成熟，未來於全球資訊網上擷取及瀏覽視訊資料將逐漸普及。視訊資料庫中的視訊資料可以藉由類似“蕃薯藤”整理網站的方式予以分類，蓋因不同類別表示不同的興趣，以便於讓使用者依據本身的需要來瀏覽擷取有興趣的視訊資料。此外，視訊資料檢索系統[1,9,10,12]的成功關鍵，在於

是否能在資料量龐大的視訊資料庫中找出使用者有興趣的視訊。因此如何歸類與整理視訊資料以便使用者查詢，是重要之研究課題。

視訊資料分類是根據視訊內容所涵蓋的主題將視訊資料歸類於已定義之分類階層中一個或多個類別中，不同的類別分別代表不同的主題，對某主題有興趣的使用者可在相關的分類中找出所需的視訊資料。然而，傳統上分類是以人工分類方式進行，既耗時又不夠客觀，分類的結果又易受分類者本身的主觀意識及過去經驗的影響而有偏差。透過電腦自動分類可以依據視訊資料內容，將視訊文件歸類至合適的類別之下。自動分類在分類過程中具有標準之規則可循，而分類的結果亦具一致性，不至於因人而異。

目前有關文件分類方法的探討，主要是以文字型態文件為主[2,5,6,7,11]，而針對視訊資料探討分類方法的研究則尚欠缺，本文以視訊資料為探討對象，研究視訊資料的分類。本文考慮以下兩項因素，以建立一個視訊資料的分類方法：

(1) 視訊資料的結構性：視訊資料的特性有異於文字，其內容因本身的結構性而形成了 sequence、scene、shot...等階層單位，故須考慮視訊內容在不同階層的代表性程度之差異。

(2) 分類架構中的特定性與涵蓋性：分類架構由具有階層關係的大小類別所構成，在經由分類程序以指定分類結果時，必須以特定性與涵蓋性兩個準則為考慮的依據。但此兩準則互相抵觸，無法同時滿足，必須在兩者之間取得平衡點。此部份主要是以文字型態文件分類方法的相關研究 [11] 為基礎。

目前尚無針對視訊資料分類之文獻，相關文獻包括文字型態文件分類之研究[2,5,6,7,11]，與視訊技術方面的研究包括視訊資料庫、以內容為基礎的視訊資料庫檢索、視訊擷取、視訊資料索引的建立、視訊資料之模式化與查詢等[1,3,4,8,9,10,12]。

視訊技術的研究主要著重在建立視訊資料的索引時給予描述，以便在檢索時根據內容進行檢索。將視訊資料片段因播放時間先後所形成的結構性階層，歸納成一由 frame, shot, scene, sequence (or segment) 等階層所組成之視訊結構，以註解或主題關鍵字來描述各階層之視訊內容包括人物、事件、地點

等，進而建立索引值來提供視訊內容查詢與擷取。例如，文獻[8]提出依據視訊查詢類別包括：視訊本身的查詢；內容主題的查詢；物件的查詢；結構性資訊的查詢等，來建立視訊資料的索引。其結構性索引之定義包括了 frame, shot, scene 與 segment。文獻[4]所提出之視訊資料模式亦定義了 compound unit, sequence, scene 與 shot 之視訊結構，並用主題索引(thematic indexing)來表示描述視訊資料內容。

文件自動分類的研究主要是探討文字型態文件之分類[2,5,6,7,11]，如文獻[11]所提出之文字型態文件自動分類的演算法:ACTION。ACTION 是依據關鍵字出現次數及分類架構中類別所在階層及類別之間的階層關係，考量特定性與涵蓋性的平衡以計算顯著值(Significance Value)。ACTION 首先建立階層式的分類架構，每個節點代表一個類別(class)。將文件中關鍵字對應於各類別的出現次數統計出後，再根據定義的分類演算法計算每一類別的顯著值，最後在顯著值較高的類別中，選取在分類架構中階層最高的類別做為分類結果。ACTION 分類演算法的目的在使分類結果於特定性(specificity)與涵蓋性(exhaustivity)兩者之間取得平衡點。特定性是指文件內容能以分類結果適切表達其主題的程度；而涵蓋性則是指文件中包含的所有內容能被分類結果涵蓋的程度[11]。在其它條件一樣的情況下，分類結果的層級愈低則其特定性較高，而涵蓋性較低，反之則異。

本文在接下來的章節中，第二節介紹視訊分類基本觀念，第三節則是討論視訊分類方法，第四節以範例分析本論文之分類方法。最後在第五節以結論做為本文之總結。

2. 視訊分類基本觀念

視訊資料分類是根據視訊內容所涵蓋的主題將視訊資料歸類於已定義之分類階層中一個或多個類別中，不同的類別分別代表不同的主題，而分類包含兩個要素：分類架構與分類演算法。本論文提出的視訊資料分類方法是以各分類類別對應於視訊內容之顯著值(Significance Value)來衡量各類別在做為分類結果的合適程度之量化指標。

(1) 建立視訊內容關鍵字時考慮不同階層視訊單元於視訊內容的代表性

由於在視訊結構中，同一個 sequence 包含數個 scenes，每個 scene 所包含的內容有其差異，在為視訊資料建立內容關鍵字時要分別針對不同視訊單元(例如：以 sequence 為視訊單元)與其中的片段(例如：以 scene 為單元)的內容指定描述詞做為關鍵字，並考慮視訊內容在不同階層的代表性，給予較高階層視訊單元的關鍵字較高的權重。因為 sequence 階層的視訊內容的描述詞目的是表達整個視訊單元最主要的內容，較不考慮細節部份，而 scene 階層的描述詞則是只描述某小段視訊片段的內容，故在意義上

sequence 階層的描述詞較 scene 階層的描述詞能代表視訊的主要內容，因此在計算顯著值時，sequence 階層關鍵字之權重要高於 scene 階層關鍵字之權重。

(2) 視訊資料分類結果兼顧特定性與涵蓋性的準則

在分類架構中愈下面層級的類別所涵蓋的範圍愈小，而一段視訊中所包含的內容又常牽涉多項主題。若指定的分類結果階層較低，則該類別能表達視訊主要內容的程度，即特定性較高；若指定的分類結果階層較高，則該分類所能涵蓋整個視訊內容的程度，即涵蓋性較高。如何在特定性與涵蓋性兩者之間取得平衡，是視訊分類需考量的重點。此部份主要是以文獻[11]之文字型態文件自動分類方法來做為視訊資料分類方法之基礎。

2.1 分類架構

在討論視訊分類的方法之前，本節先討論分類架構。依據已定義之分類架構，來決定將視訊資料歸類至分類架構的一個或多個類別。分類架構是由類別與子類別之間形成的階層式架構，類別下包含數個子類別，而子類別下又可分為更小的子類別，依此類推。

分類架構 C 為一有序對： (C_i, l_i) ， $1 \leq i \leq n$ ，其中 C_i 是分類架構 C 中已定義的類別， l_i 是 C_i 的階層數(level number)， n 是分類架構 C 中所有類別的數目。以樹狀結構來描述分類架構所包含的類別與子類別及其間的階層關係，其中每個 tree 為以 level 0 的節點為根節點的樹狀資料結構，每一節點 v 對應分類架構 C 之一序對 (C_i, l_i) ，節點 v 代表 C_i 類別，節點 v 的 level number (level(v)) 為 l_i 。

2.2 視訊結構與關鍵字

在進行文字型態文件的分類之前，必須先對文件進行前置處理，擷取文件中具代表性、可以表達內容的關鍵字。但視訊內容的關鍵字描述則無法直接擷取，雖然已有許多研究探討視訊物件自動辨識與分析[3,12]等前置處理以進行視訊內容之建置，但擷取視訊的關鍵字的過程不在本論文討論的範圍之內，本論文假設視訊內容之關鍵字描述已經前置處理好，係經由人工加以指定輸入建置[8]或經由視訊物件自動辨識分析前置處理[3,12]。視訊資料之結構性可分為 compound unit、sequence、scene、shot 等階層[4]，定義如下：shot 由一個或多個 frames 所組成，多個 shots 組成一個 scene，多個連貫有意義的 scenes 組成一個 sequence (或叫做 segment)，相關的 sequences 則可組成一 compound unit，多個 compound units 可再組成一 compound unit，形成一遞回(recursive)定義之關係。依據所需分析之資訊內容的詳細程度，可選定至 scene 或是 shot 階層。

例如以 sequence 視訊單元代表一個新聞事件報導單元，而其中的新聞片段為 scene。針對 sequence 與 scene 等視訊結構階層，對視訊內容指定描述詞或

註解做為關鍵字，包括人物、事件、地點等。圖 1 以劉邦友官邸血案報導之視訊檔案為例，為簡化說明，假設此視訊檔案只包含一 sequence。此視訊資料中所包含的 sequence 與 scene 的關鍵字如下，括號內為該關鍵字在描述詞中的出現次數：

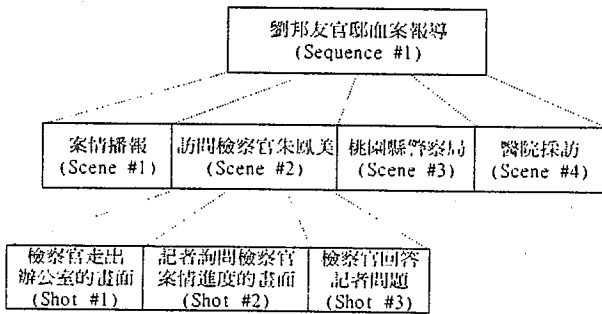


圖 1：視訊結構範例

sequence #1

劉邦友、血案(2)、官邸(1)、槍殺、桃園縣、朱鳳美、警察局、鄧文昌、案情

scene #1

劉邦友、血案(1)、朱鳳美、桃園縣、警察局、鄧文昌

scene #2

官邸(3)、血案(2)、朱鳳美、祕密證人、案情、暗場、警衛、記者會、投案

scene #3

劉邦友、警察局、證據、作案、槍、祕密證人、目擊證人、歹徒、指紋

scene #4

醫院、加護病房、中彈、彈殼、手術、縣議員、病情

3. 視訊分類方法

本節將本文的視訊資料分類方法分為兩部份，分別在 3.1 及 3.2 小節探討。3.1 小節針對視訊資料的特性，以及關鍵字對應到類別的彈性，分別設定關鍵字視訊權重與關鍵字類別權重兩種權重值，以便求出各類別本身的視訊分類值，視訊分類值較高的類別較接近視訊內容，但尚不考慮類別之間的階層關係。3.2 小節以參考文獻[11]中針對文字型態文件的分類演算法為基礎來設計本論文視訊資料分類演算法，考慮類別之間的階層關係，並以特定性(specificity)與涵蓋性(exhaustivity)兩個準則求出各類別的顯著值，做為最後決定分類結果的依據。

3.1 視訊分類—依據視訊分類值

本小節目的在為每個類別分別計算其視訊分類值，暫不考慮類別之間的階層關係。單就視訊分類值而言，其值較高的類別較接近視訊內容，但真正決定分類結果的因素則是取決於 3.2. 小節的顯著值。

3.1.1 關鍵字視訊權重

本文在視訊資料的關鍵字定義上，是以經前置處理建置對視訊資料各階層內容包括 sequence, scene 等階層視訊片段內容之註解或描述詞做為關鍵字，針對視訊資料結構的特性，為衡量各關鍵字於表達視訊主題的重要程度，本文定義了一個權重：關鍵字視訊權重 K_{VW}，當 K_{VW} 值愈高，則表示該關鍵字在表達視訊主題的重要性愈高。定義 1 以 K_{VW}(k_p) 表示關鍵字 k_p 在視訊內容中的重要程度，若 K_{VW}(k_p) > K_{VW}(k_q), k_p ≠ k_q，則表示關鍵字 k_p 在視訊中的重要性高於關鍵字 k_q。

定義 1：關鍵字視訊權重(Keyword Video Weight)，以 K_{VW}(k_p) 表示，定義為關鍵字 k_p 在表達視訊主題的重要程度，其算式如下：

$$K_{VW}(k_p) = \sum_l Occur(k_p, l_i) * (h - l_i) \quad l_i = 0 \dots h-1$$

h：視訊階層的高度，l_i：視訊階層數的值

$$Occur(k_p, l_i) = \sum_{A_j} freq(k_p, A_j)$$

A_j 為在視訊階層 l_i 中之視訊單元；Occur(k_p, l_i) 為關鍵字 k_p 在視訊階層 l_i 中的描述詞的出現次數，freq(k_p, A_j) 則為 k_p 在視訊單元 A_j 中的描述詞的出現次數。計算 K_{VW}(k_p) 的方式，是針對視訊中的關鍵字 k_p 分別計算其在每個視訊階層中描述詞出現的次數 Occur(k_p, l_i)，並乘上 (h - l_i)，最後再將各階層的乘積加總，所得結果就是關鍵字 k_p 的關鍵字視訊權重。若關鍵字在 sequence 與在 scene 階層出現的次數一樣，因 sequence 的視訊階層數 i 較小，所乘上的 (h - l_i) 較大，乘積會比在 scene 階層的結果高，定義 1 給予在 sequence 階層的關鍵字比在 scene 階層的關鍵字較高之視訊權重。

視訊資料由大到小可分為 sequence、scene、shot... 等階層，一段視訊單元 sequence，包含了許多不同場景所形成的 scene。每一個 sequence 和其中的每一個 scene 皆分別有描述其內容的關鍵字，雖然每個 scene 中所包含的人、物、主題內容各有不同，但就整個 sequence 而言其主要重點則是反應在 sequence 的關鍵字中。因為 sequence 較 scene 的階層為高，涵蓋的視訊長度較長，sequence 較 scene 具有整段視訊內容的代表性，理論上在 sequence 階層定義的關鍵字較在 scene 階層定義的關鍵字代表性為高，故 sequence 階層的關鍵字應給予較高的權重。例如，圖 2 之視訊檔案包含 sequence 與 scene 兩階層，分別給予不同的視訊階層數值。

level 0	sequence #1 劉邦友官邸血案報導			
level 1	scene #1 案情播報	scene #2 訪問檢察官 朱鳳美	scene #3 桃園縣警察 局偵辦狀況	scene #4 醫院採訪

圖 2：視訊中不同結構階層的視訊階層數

K_{VW}(k_p) 之計算：本例而言 h=2 (即 sequence

與 scene 兩階層)，sequence 的階層數為 0，scene 的階層數為 1。以 2.2 小節中的關鍵字“血案”為例，出現在 sequence 階層中的次數為 2，在 scene #1 的出現次數為 1，在 scene #2 的出現次數為 2，則該關鍵字的視訊權重為： $KVW(\text{“血案”}) = 2 * (2-0) + (1+2) * (2-1) = 7$ ，而“官邸”的視訊權重為： $KVW(\text{“官邸”}) = 1 * (2-0) + 3 * (2-1) = 5$ 。因為 $KVW(\text{“血案”})$ 大於 $KVW(\text{“官邸”})$ ，可看出“血案”較“官邸”更能呈現視訊的內容。

3.1.2 關鍵字相關類別權重

在為每個關鍵字計算出其關鍵字視訊權重，以衡量其在視訊中的重要性後，接下來則是將關鍵字對應到與其相關的類別。定義 2 定義了關鍵字相關類別權重(Keyword Class Weight)， $KCW(k_p, C_i)$ ，為關鍵字 k_p 對應至類別 C_i 的權重值。權重值愈大，則表示該關鍵字歸類在該類別的可能性愈大。

定義 2： 關鍵字相關類別權重(Keyword Class Weight)，以 $KCW(k_p, C_i)$ 來表示，定義為關鍵字 k_p 對應至類別 C_i 的權重值。

理論上當某個關鍵字出現時，可能對應到的類別不只一個，且對應的機率有其差異。透過視訊內容關鍵字對應至分類架構中各類別的權重設定值的差異，避免將關鍵字完全對應至單一類別。以“血案”這個關鍵字為例，可能對應到“政治”、“社會”、“刑事案件”等類別之新聞，但機率值有差異，假設分別為 0.15、0.25、0.60。為了查詢某個關鍵字對應至某類別的的權重值，在此假設已經根據過去統計資料為每個類別建立其相關字彙庫。每個類別 C_i 的字彙庫中包含了與該類別相關的關鍵字 k_p 與其對應至該類別的權重。以“血案”與“官邸”兩關鍵字為例，對應至相關類別的 KCW 如下： $KCW(\text{“血案”}, \text{“政治”}) = 0.15$ ， $KCW(\text{“血案”}, \text{“社會”}) = 0.25$ ， $KCW(\text{“血案”}, \text{“刑事案件”}) = 0.6$ ， $KCW(\text{“官邸”}, \text{“政治”}) = 0.8$ ， $KCW(\text{“官邸”}, \text{“社會”}) = 0.2$ 。 $KCW(k_p, C_i)$ 代表關鍵字 k_p 對應至相關類別 C_i 的權重，若 $KCW(k_p, C_i) > KCW(k_p, C_j)$ ，其中 $i \neq j$ ，則代表當關鍵字 k_p 出現時，屬於 C_i 類別的權重較屬於 C_j 類別的權重高。例如：當“血案”關鍵字出現時，屬於“刑事案件”的權重(0.6)較屬於“政治”新聞的權重(0.15)高。

3.1.3 依關鍵字歸屬類別值

定義 3： 依關鍵字歸屬類別值(Keyword Classification Value)，以 $KCV(k_p, C_i)$ 來表示，定義為關鍵字 k_p 對於視訊資料歸屬於類別 C_i 的貢獻程度。

$$KCV(k_p, C_i) = KVW(k_p) * KCW(k_p, C_i)$$

在求出視訊中每一個關鍵字的視訊權重與該關鍵字對應到類別的權重之後，將兩者相乘即得到

KCV 。 $KCV(k_p, C_i)$ 的意義為就關鍵字 k_p 而言，對於視訊資料歸屬於類別 C_i 的貢獻程度。舉例來說：3.1.1 小節已算出“血案”和“官邸”兩個關鍵字的視訊權重如下： $KVW(\text{“血案”}) = 7$ ， $KVW(\text{“官邸”}) = 5$ ，為求出此兩個關鍵字對應至“政治”類別的 KCV ，算式如下： $KCV(\text{“血案”}, \text{“政治”}) = KVW(\text{“血案”}) * KCW(\text{“血案”}, \text{“政治”}) = 7 * 0.15 = 1.05$ ； $KCV(\text{“官邸”}, \text{“政治”}) = KVW(\text{“官邸”}) * KCW(\text{“官邸”}, \text{“政治”}) = 5 * 0.8 = 4$ ，以“官邸”對應至“政治”類別的 KCV 值為大。意義為單就“政治”類別而言，關鍵字“官邸”較關鍵字“血案”對於以“政治”類別做為分類結果的貢獻程度較高。

3.1.4 視訊分類值

定義 4： 視訊分類值(Video Classification Value)，以 $VCV(C_i)$ 來表示，定義為視訊資料歸屬於類別 C_i 的適度性，其算式如下：

$$\begin{aligned} VCV(C_i) &= \sum_{p=1..m} KCV(k_p, C_i) \\ &= \sum_{p=1..m} (KVW(k_p) * KCW(k_p, C_i)) \end{aligned}$$

$p = 1, 2, \dots, m$ ； m 是視訊資料的所有關鍵字的個數

計算視訊資料中的每個關鍵字對應至每個類別的 KCV 後，可求出每個類別的視訊分類值。任一類別 C_i 的視訊分類值是所有關鍵字對應到該類別的 KCV 之和。類別 C_i 的視訊分類值 $VCV(C_i)$ 與 $\sum KCV(k_p, C_i)$ 呈正比。 $\sum KCV(k_p, C_i)$ 的意義為就所有的關鍵字而言，對於類別 C_i 的視訊分類值的貢獻程度總和，其值愈高表示該類別較適合做為分類結果。

假設視訊中只有“血案”與“官邸”兩個關鍵字，分別計算“政治”、“社會”、“刑事案件”類別的視訊分類值如下： $VCV(\text{“政治”}) = KVW(\text{“血案”}) * KCW(\text{“血案”}, \text{“政治”}) + KVW(\text{“官邸”}) * KCW(\text{“官邸”}, \text{“政治”}) = 7 * 0.15 + 5 * 0.8 = 5.05$ ； $VCV(\text{“社會”}) = KVW(\text{“血案”}) * KCW(\text{“血案”}, \text{“社會”}) + KVW(\text{“官邸”}) * KCW(\text{“官邸”}, \text{“社會”}) = 7 * 0.25 + 5 * 0.2 = 2.75$ ； $VCV(\text{“刑事案件”}) = KVW(\text{“血案”}) * KCW(\text{“血案”}, \text{“刑事案件”}) = 7 * 0.6 = 4.2$ 。視訊分類值較大的類別較其它類別更能視為視訊所屬的類別。該視訊屬於“政治”類的合適程度最高，其次為“刑事案件”類，最後為“社會”類。

3.2 視訊分類—考量特定性與涵蓋性

本小節計算類別的顯著值 (Significance Value)，以做為衡量各類別做為分類結果的合適程度的量化指標。顯著值係以 3.1 小節計算的視訊分類值為基礎，考慮了類別之間的階層關係，並設法取得特定性與涵蓋性的平衡，最後再選擇顯著值最高的類別做為分類結果。此小節主要是以文獻[11]的文字文

件分類演算法中關於特定性、涵蓋性與顯著值計算之分類法則，來做為視訊分類方法之基礎。所參考的主要法則如下（修改自[11]）：

法則1：對於任何視訊資料檔案A及分類架構C，若存在任意兩類別具有相同的視訊分類值（Video Classification Value），則在分類架構階層數較大的類別應給予較高的顯著值。

若兩類別有相同的視訊分類值，則必須考慮特定性。階層數較大的類別較能呈現文件內容的深度，亦即特定性較高，故應在計算顯著值時，給予較大的值。

法則2：對於任何視訊資料檔案A及分類架構C，若存在超過一個以上的類別具有相同的顯著值，則選擇階層數最小的一個或數個類別做為分類結果。

顯著值是以各類別做為分類結果的合適程度的衡量指標，如果兩個結點(即類別)有相同的顯著值，則有相同的合適性。此時，必須考慮涵蓋性，階層數較小的類別較能包含文件內容的廣度，故以它為最後的分類結果。依據上述法則計算顯著值時，在分類架構中，愈下面階層的類別，即階層數較大的類別，在概念上較能表達視訊內容的特定性，因此節點v的特定性可用節點的階層數（level(v)）來表示其特定程度。參考自文獻[11]文字文件分類有關顯著值計算之定義，本文新定義一涵蓋視訊分類值於定義5，並重新定義視訊分類顯著值之計算於定義6。

定義5：以節點v為根節點的子樹(subtree)的涵蓋視訊分類值(Exhaustive Video Classification Value)，以EVCV(v)表示，定義為節點v本身及子樹中所有successor節點的視訊分類值總和。

$$EVCV(v) = \sum VCV(j) ; j = v \text{ or } j \text{ is a successor of } v$$

EVCV(v)可以視為以v為根節點的子樹中所有不同階層的節點在表達視訊內容的整體涵蓋概念上的和，當v為終端節點(leaf node)時，則EVCV(v) = VCV(v)。

定義6：顯著值(Significance Value)，以SV(v)表示，定義為以v節點為根節點的子樹的涵蓋視訊分類值EVCV(v)與v節點的階層數的乘積，及v節點的所有ancestor節點的視訊分類值與所在階層數乘積之總和。

$$SV(v) = EVCV(v) * level(v) + \sum_j (VCV(j) * level(j))$$

$$= \sum_i (VCV(i) * level(v)) + \sum_j (VCV(j) * level(j))$$

$i = v \text{ or } i \text{ is a successor of } v, j \text{ is an ancestor of } v$

當v為終端節點時，SV(v)等於節點v到根節點的路徑上所有節點的視訊分類值與其所在的階層數分別相乘後的乘積總和。SV(v)包含兩個項目：

(1)其中EVCV(v) * level(v)可以分解為 $\sum (VCV(i) * level(v))$ ，意義為以v為根節點的子樹中所有節點i(包含v節點本身)的視訊分類值與v節點階層數的乘積之和。注意所乘的皆為v節點的階層數而非節點i本身的階層數。此項目包括EVCV以計算涵蓋性對顯著值之貢獻程度。

(2) $\sum (VCV(j) * level(j))$ 則是節點v的所有ancestor node的視訊分類值分別乘上其所在階層數的乘積之和。此項目包括從根節點到v節點之路徑中，所有節點之VCV(j) * (level(j))，反應特定性對顯著值之貢獻程度。

在計算出所有節點(類別)的顯著值後，選擇顯著值最高的類別做為最後的分類結果，若有一個以上的節點同時具有最高顯著值，則選擇階層最高者。若欲將視訊資料歸類於t個類別，則可選擇顯著值最高的前t個類別作為分類結果。

4. 範例分析

本節以範例說明如何利用本研究的視訊分類方法，計算各類別的顯著值，求出合適的分類結果。假設已經根據視訊的關鍵字計算出每個類別的視訊分類值，至於視訊分類值的計算方式如3.1節所述，不再贅述。本節則是說明由視訊分類值進一步求出顯著值的過程。表格1是各類別在分類架構中的階層數與該類別的視訊分類值：

表格1：類別階層數與視訊分類值

類別名稱	類別所在階層數	視訊分類值	類別名稱	類別所在階層數	視訊分類值
新聞	0	0	縱火	3	0
社會	1	1	凶殺	3	5
環保	2	0	走私	3	0
刑事案件	2	2	地方建設	2	1
賭博	3	1	宗教	2	0
強盜	3	0			

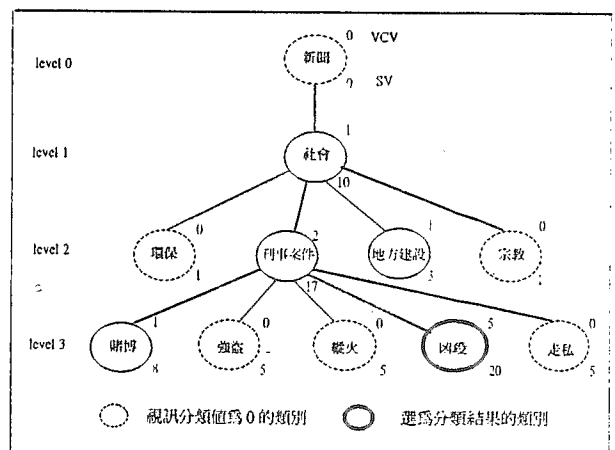


圖3：分類架構

此視訊資料對應的分類架構如圖 3，圖中每個節點代表一個類別，類別的視訊分類值標於節點的上方，而顯著值則標於節點的下方。顯著值的計算過程則列於表格 2 中。

表格 2：顯著值的計算過程

類別名稱	階層數	顯著值	顯著值的計算過程
社會	1	10	$EVCV=1+2+1+1+5=10$ $SV=10*1+0*0=10$
環保	2	1	$SV=0*2+1*1+0*0=1$
刑事案件	2	17	$EVCV=2+1+5=8$ $SV=8*2+1*1+0*0=17$
賭博	3	8	$SV=1*3+2*2+1*1+0*0=8$
強盜	3	5	$SV=0*3+2*2+1*1+0*0=5$
縱火	3	5	$SV=0*3+2*2+1*1+0*0=5$
凶殺	3	20	$SV=5*3+2*2+1*1+0*0=20$
走私	3	5	$SV=0*3+2*2+1*1+0*0=5$
地方建設	2	3	$SV=1*2+1*1+0*0=3$
宗教	2	1	$SV=0*2+1*1+0*0=1$

結果分析：

1. “凶殺”類別具有最高的視訊分類值($VCV = 5$)，同時也是在最下方的階層中，因此其能代表視訊內容的特定性亦最高。由表格 2 得知，“凶殺”具有最高的顯著值($SV=20$)，因此被選為分類結果。
2. “刑事案件”類別本身的視訊分類值並不是最高，但因其下的兩個子節點“賭博”與“凶殺”的視訊分類值都大於 0，有助於涵蓋視訊分類值提高，同時影響了顯著值($SV=17$)。由表格 2 可知其顯著值為所有節點的次高者。對於非終端節點而言，除了本身的視訊分類值之外，在以它為根節點的子樹中其它節點的視訊分類值亦是會影響其顯著值的因素。因為在涵蓋視訊分類值的計算中，是將子樹中所有節點的視訊分類值加總，代表整個子樹中所有主題近似但階層不同的類別的整體概念之和。

5. 結論

視訊資料庫的發展以及 WWW 的使用日趨普及，如何將視訊資料分門別類以協助使用者瀏覽搜尋其有興趣的視訊資料，是重要的研究課題，本研究即是探討視訊資料之分類。本研究之成果可歸納如下：

1. 本文是依據視訊資料內容的結構性來進行視訊資料分類。視訊資料的特性有異於文字，其內容因本身的結構性而形成了 sequence、scene 等階層單位，所提出的方法考量視訊內容在不同階層的代表性程度之差異，將不同階層關鍵字相對於視訊的重要性給予不同之權重值。
2. 本文參考文獻[11]文字文件分類的演算法，針對視訊資料的特性予以修改，提出視訊資料分類的程序，以滿足特定性與涵蓋性為著眼點。本文提出之方法，使視訊分類的過程有規則可循，避免由人進行分類造成分類結果不一致的情況。

參考文獻

- [1] M. G. Brown, J. T. Foote, G. J. F. Jones, K. Sparck Jones, S. J. Young, “Automatic Content-Based Retrieval of Broadcast News”, Proc. of the ACM Intl. Conf. on Multimedia, San Francisco, CA, 1995
- [2] Valery I. Frants and Nick I. Kamenoff, “One Approach to Classification of Users and Automatic Clustering of Documents”, Information Processing and Management, Vol. 29, No. 2, pp. 187-195, 1993
- [3] Borko Furht, Stephen W. Smoliar and HongJiang Zhang, *Video and Image Processing in Multimedia Systems*, published by Kluwer Academic Publishers, 1995
- [4] Rune Hjelsvold and Roger Midstraum, “Modeling and Querying Video Data”, Proc. of the 20th VLDB Conference, Santiago, Chile, 1994
- [5] David D. Lewis, “Evaluating and Optimizing Autonomous Text Classification Systems”, Proc. of the 18th ACM SIGIR Conf. On Research and Development in Information Retrieval, Seattle Washington, 1995
- [6] C. T. Meadow, *Text Information Retrieval Systems*, Academic Press, 1992
- [7] Susanne Oranger, “The Newspaper Image Database: Empirical Supported Analysis of User's Typology and Word Association Clusters”, Proc. of the 18th ACM SIGIR Conf. On Research and Development in Information Retrieval, Seattle Washington, 199
- [8] L. A. Rowe, J. S. Boreczky, and C. A. Eads, “Indexes for User Access to Large Video Databases”, Storage and Retrieval for Image and Video Databases II, IS&T/SPIE Symp. On Elec. Imaging Sci. & Tech., San Jose, CA, February 1994
- [9] Stephen W. Smoliar and H. J. Zhang, “Content-based Video Indexing and Retrieval”, IEEE Multimedia, Summer 1994
- [10] J.K. Wu, A. Desai Narasimhalu, B.M. Mehtre, C.P. Lam, Y.J. Gao, “CORE: a Content-based Retrieval Engine for Multimedia Information Systems”, Multimedia Systems, Vol. 3, pp. 25-41, 1995
- [11] Jacqueline W.T. Wong, W.K. Kan, Gilbert Young, “Action: Automatic Classification for Full-Text Documents”, SIGIR FORUM, ACM Special Interest Group on Information Retrieval, Vol. 30, No. 1, pp.26-41, Spring 1996
- [12] H. J. Zhang, C. Y. Low, S. W. Smoliar and J. H. Wu, “Video Parsing, Retrieval and Browsing: An Integrated and Content-Based Solution”, Proc. of the ACM Intl. Conf. on Multimedia, San Francisco, CA, 1995