

Objectionable Streaming Video Filtering Based on Video Classification

基於圖片分類之不當串流影片過濾

Yi-Leh Wu (吳怡樂)

國立台灣科技大學資訊工程學系

Email: ywu@csie.ntust.edu.tw

Maw-Kae Hor (何瑁鎧)

國立政治大學資訊科學系

Email: hor@cs.nccu.edu.tw

Cheng-Yuan Tang (唐政元)

華梵大學資訊管理學系

Email: cytang@cc.hfu.edu.tw

Xing-Hong Ye (葉星宏)

華梵大學資訊管理學系

Email: m9625016@cat.hfu.edu.tw

摘要：由於電腦的普及與網路共享的發達，一般人越來越容易藉由網路接觸到令人反感的資訊，其實令人反感的資訊並不適合尚未成年者觀看，也會對未成年者的心靈產生傷害，所以如何杜絕或防範未成年經由網路取得這些不當資訊，已經成為當前的重要課題。本研究對於不當資訊的定義為色情影片，內容多為限制級與不正確的性觀念。

不當影片中通常會包含正當影片片段與不當影片片段，影片內也會出現較長時間的不當資訊影片片段，本文探討在出現多少時間的不當資訊影片片段後我們便可以認定此影片為不當影片。

在我們的研究中提出以影片片段方式觀測是否為不當影片片段，並使用兩層式過濾器來分類不當影片。在第一層使用的是傳統的圖片過濾器對影片片段中擷取出的影像進行過濾，得到一系列圖片分類的結果，分析一系列圖片過濾器的結果並產生特徵，利用這些特徵，透過第二層的支持向量機分類不當影片片段，並觀察影片中色情程度片長時間達某一門檻進行影片分類，已達到過濾不當影片資訊。

藉由本文提出的方法，降低圖片過濾器的誤判率，提昇對於不當影片過濾的準確率。

關鍵詞：不當影片偵測、不當影片分類、圖片過濾器、特徵擷取、支持向量機

Abstract— Due to the popularity of the computer-sharing networks, the general public has more easy access to objectionable Internet information, which is not suitable for minors. Therefore, the methods to stop and prevent the minors to access these improper information on the Internet become an important issue. In this work, the improper information is defined as the pornographic images and video streams.

However, the improper video streams usually contain both benign and objectionable video segments. We propose a two-tier SVM classifier method to decide whether a given video stream is a proper or an improper video film and the shortest time elapse required to determine. In the first tier, we adopt the traditional image filter to classify video frames and get a series image classification result, which may contain false-positive and/or false-negative classifications. From the first tier image classification result we then extract the potential video features. In the second tier, we use a second

SVM classifier to classify video clips and decide whether the given video stream is benign or objectionable in the shortest time frame possible.

We show that the propose methods can effectively compensate the inaccuracy caused by using the image classification results alone to classify videos and to enhance the accuracy of video stream classification.

Keyword— objectionable video detection, objectionable video classification, image filter, feature detection, support vector machine.

Keyword : objectionable video detection, objectionable video classification, image filter, feature detection, support vector machine

一、前言

目前的網路傳輸速度快，電腦設備越來越好，多媒體影音在網際網路上呈現的不只是存取圖片，取而代之的是影音，以現今較有名的網路影音共享平台 Youtube 來說，只要鍵入關鍵字搜尋，選擇想要觀賞的影片只需要幾秒鐘的時間即可，這類型的網路共享平台越來越多，也包含了不當影片共享平台，如何防範不當資訊隨處可得，就顯的十分的重要。

Wu 等人[2]，提出了針對影像是否為不當資訊的過濾器，輸入一張影像經由影像過濾器會產生一個分數，分數介於 $[0, 2]$ 之間視為有較高的機率是屬於不當影像，分數介於 $[-2, 0]$ 之間視為有較高的機率是屬於正當影像，分數為 -10 表示有非常高的機率是屬於正當影像。

林柏璋[1]，依據影像過濾器，針對整部正當影片與不當影片中分數的分佈，提出了幾個方法來分析影像過濾器的結果並產生特徵，利用獲取的特徵透過支持向量機分類不當影片。

雖然林柏璋[1]針對正當影片與不當影片過濾高達 96.27% 的正確率，但前提是必須把整部影片作過濾才能判斷出是否為不當影片，本文主要目的是希望過濾片段影片資訊，在影片播放到某個時間點就能夠判斷出是否為不當影片，已達到過濾串流影片的效果。

二、圖片過濾器

本文基於[2]的影像過濾器作為影像分類的依據，輸入一張影像經由影像過濾器得到影像的分數，假如分數為正數值，則此張分類為不當影像，且數值越大代表輸入的影像為不當影像的機率就越高，反之分數為負數值，則分類為正當影像，且數值越小代表輸入的影像為正當影像的機率就越高。影像過濾器的正數值範圍是 $[0, 2]$ ，負數值範圍是 $[-2, 0]$ ，假如輸入的影像經由影像過濾器計算出來的分數為 -10 ，則表示此影像

有非常非常高的機率為正當影像。

我們從影片擷取出的影像，以人工標記方式紀錄影像是否為正當影像或不當影像，並將人工標記結果與影像過濾器計算影像分數的結果作比較，在不當影片中使用影像過濾器的正確率為 55.60%，在正當影片中使用影像過濾器的正確率為 88.31%，其總準確率為 71.96%，由結果來說明其影像過濾器對於正當影像的過濾準確率優於過濾器對於不當影像的過濾準確率。

影像過濾器無法較準確的過濾出不當影像的原因是因為人工標記的依據是檢查影像中是否有出現人體隱私的部份(如：露點、生殖器官 …等)，而影像過濾器是依據色彩的分佈來計算影像的分數(如圖 1、圖 2、圖 3)。假設影像中有較多的顏色與膚色相似，會造成影像過濾器物判為不當影像，如：影像中有很大的大臉，人體近拍特寫…等狀況發生時，假使這些情況都不是含有不當影像的資訊，也會使得影像過濾器誤判成不當影像；影像中的光線條件也會使的影像過濾器的誤判，假如光線偏暗或偏亮，會造成原本是不當影像被分類為正當影像；或者影片本身有經由壓縮，使的擷取出來的影像也會因為失真而造成影像過濾器的誤判。



圖 1 圖片過濾器 I/O 示意圖 1



圖 2 圖片過濾器 I/O 示意圖 2



圖 3 圖片過濾器 I/O 示意圖 3

三、支持向量機

支持向量機(support vector machine)是最近幾年常被使用到的方法[4]，並且廣泛的應用在各個領域上。例如資料探勘、影像處理等等。

3.1 SVM

[5]假設空間 R^d 中存在著屬於不同的兩群組資料點，我們想找出可以區別兩群的最佳 Hyper-plane H (實線)，利用 SVM 可根據資料的特性將空間 R^d 的資料點作區別，如圖 4。

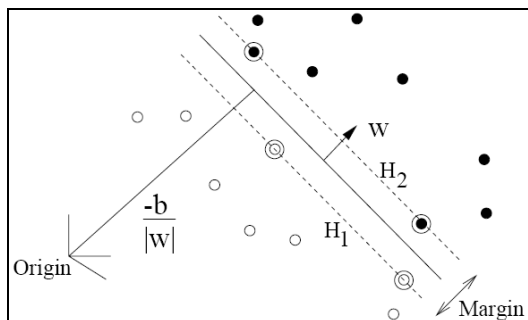


圖 4: Hyper-plane 示意圖

利用已知 Training Data Sets : $\{x_i, y_i\}$,

$i = 1, 2, \dots, n$, $x_i \in R^d$, $y_i \in \{1, -1\}$, 希望利用

Training Data 找出一最佳 Hyper-plane H , 以利將未知的 x_i 歸類。

3.2 LIBSVM : A Library for Support Vector Machines

LIBSVM[6][7]是一個 Open Source 支援多種平台及程式語言，它是一套支援向量機軟體，提供多種決策核心(kernel)：Classification (C-SVC、nu-SVC)、Regression (epsilon-SVR、nu-SVR)、Distribution Estimation (one-class SVM)，同時也支援多目標分類。

LIBSVM 內包含了三支主要的程式：svm-scale、svm-train、svm-predict。

svm-scale：將原始資料內容做正規化處理。

svm-train：將整理成 LIBSVM 格式的 training data 做訓練，訓練後可以得到一個 model，依靠計算求得的 model 來做影片片段預測。

svm-predict：依據訓練得來的 model 進行 test data 預測。

3.3 LIBSVM file format

我們要使用 libsvm 之前，我們必須要先將蒐集到的資料整理成 libsvm 所能

支援的格式。本文使用不當影片片段的 label 為 1，正當影片片段的 label 為 -1，作為 LIBSVM 的資料標記。

表 1 SVM 資料格式

Label	Index: Value	...	Index: Value
Label	Index: Value	...	Index: Value

1. Label：使用 1 或 -1 區別是否為該類。

2. Index：有順序的索引。

3. Value：使用的特徵值。

3.4 Arguments

在 training 時，我們常常會使用下列三個比較重要的參數：

1. gamma：set gamma in kernel function (default

1/k, k: 輸入 資料筆數)。

2. cost : set the parameter C of C-SVC , epsilon-SVR , 和 nu-SVR (default 1)。
3. cross validation : 隨機亂數將 training data 分成 n 組來計算 cross validation 來測試我們所找到的 gamma、cost 是否為特定輸入時的好參數。

四、特徵擷取與標記類別方式

本論文是以 sliding window 方式提出了一套方法，利用圖片過濾器[2]計算從影片擷取出來的串流圖片，並得到一系列的分數，以 window 為單位(本文以 20 張為影片片段的最小單位)的偵測方式來取代傳統單張影像的偵測，藉由連續影像間分數的關係來判定其影片片段是否為正當影片片段或不當影片片段。

Step1: 我們訂立了一個固定大小的影片片段長度 (20 張/1 分鐘)，首先擷取影片初始的第一個影片片段，並觀察影片片段內擷取出來的影像是屬於正當圖片或不當圖片，影片片段內的正當圖片數量大於不當圖片數量時，視影片片段為正當影片片段，並繼續以一個固定大小的影片片段對下一時間點的影片片段進行正當與不當圖片數量的判斷。

Step2: 當某一個影片片段內的不當圖片多於正當圖片時，視為不當影片片段，接著會以原有的影片片段資訊在加入新的串流影像進來，而影片片段便會以原時間長度在增加一單位長度，並重新判斷影片片段內的正當與不當圖片數量是否為正當或不當影片片段，假如影片片段為不當影片片段，則繼續增加新的串流影像來判斷，此時會使的不當影片片段的時間長度變得越來越長。

Step3: 而不當影片片段會隨著時間長度慢慢變

長，直到影片片段內開始含有正當影片的資訊，慢慢稀釋掉影片片段內原有的不當資訊含量，使的影片片段又回歸到正當影片片段，此時，系統便會以原始大小的影片片段時間偵測下一筆影片片段，而再次判斷是以 Step1 進行偵測還是以 Step2 進行偵測，直到影片結束為止。

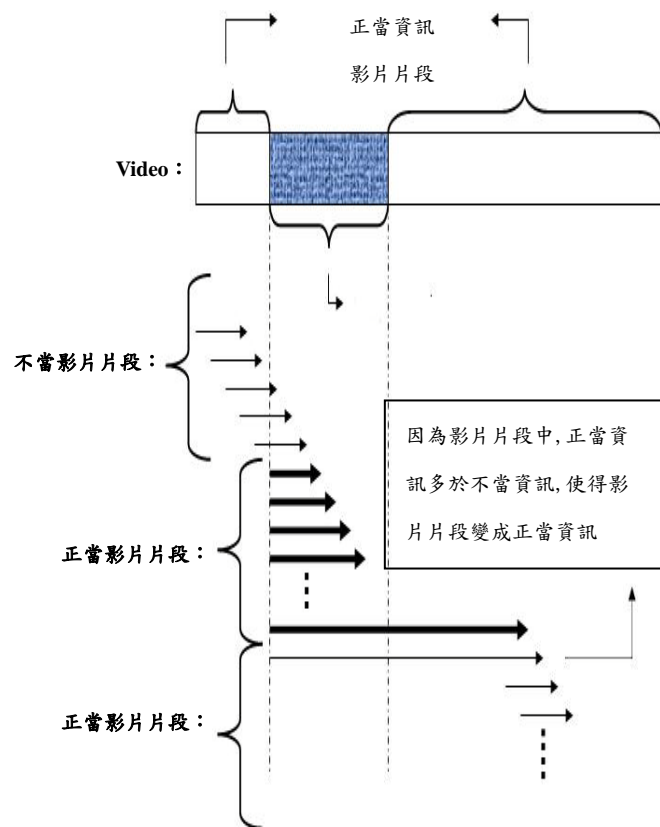


圖 5 影片片段偵測演算法

針對以上的規則我們可以得到影片片段標記為正當或不當影片片段。我們由不當影片擷取出的不當影片片段給予標記 1, 由正當影片擷取出的正當影片片段給予標記-1, 作為標記的定義。

4.1. 值方圖

我們設定了 8 個分數範圍：(1, 2], (0.5, 1], (0, 0.5], (-1, 0], (-2, -1], (0, 2], (-2, 0] 和 -10, 計算影片片段中影像分數的直方圖分佈。

以正當影片來觀察分數的分佈(如圖 6), 發

現分數大多分佈在小於 0 和 -10 之間，只有少部份的影像分數大於 0。

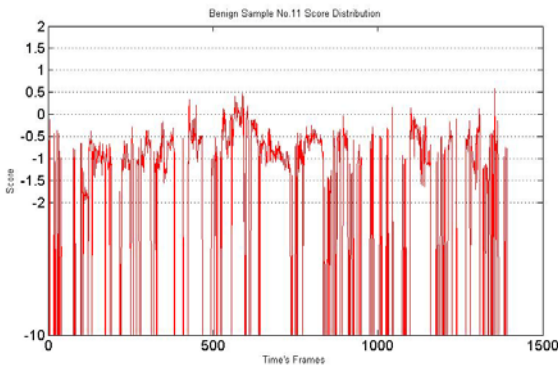


圖 6 正當影片分數分佈

以不當影片來觀察分數的分佈(如圖 7)，發現分數有較多的分佈座落在正值，有較多的分佈在 -0.5 到 1.5 之間，只有少部份的影像分數小於 -0.5 和 -10。

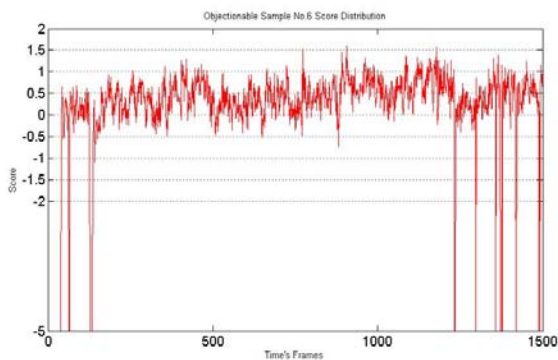


圖 7 不當影片分數分佈

由以上範例可知，使用 8 個分數範圍當做分類器的特徵，是有助於我們分類影片片段是屬於正當影片片段與不當影片片段。

4.2 平均值與標準差

平均值可以得知影片片段中的分數是偏向負值或正值，是一個有效區別正當影片片段與不當影片片段的特徵；而標準差可以得知影片片段內分數間的離散程度。

4.3 正值相鄰連續特性

我們觀察影片片段中正值分數相鄰的數量，希望以此關係幫助判斷當有少量分數誤判為極值(-10)時，影片片段不會導致不當影片片段判斷成正當影片片段。

綜觀以上的幾個分數分佈的情況，我們集結了 13 個特徵當作屬性，對每個偵測出來的影片片段計算 13 種屬性的程度值，當作 SVM 的屬性作測試與預測（如表 2）。

表 2 特徵擷取資訊

特徵 (屬性)	特徵數
1. 8-Histogram	8
2. Window 平均數	1
3. Window 標準差	1
4. Window 變異數	1
5. 正數連續性	1
6. 正數最大連續性	1
總和	13

五、實驗部份

本文提出了以影片片段方式偵測影片的方法探討影片中的色情程度達到多少時間變能夠有較高的分類影片準確率，影片片段偵測方法結合了圖片過濾器、影片片段偵測法與 SVM 的整合，針對影片片段做偵測。

5.1 影片片段偵測結果

我們收集了 75 部正當影片與 75 部不當影片，各取 50 部作 training data，剩下各 25 部當作 testing data，正當影片種類包含連續劇，影集與電影，多為東方人角色居多，包含有碼無碼，長片與短片等影片種類。

在 training 部份我們從 50 部正當影片與 50 部不當影片隨機各取出 55000 筆影片片段給予標記當作 training data，在 testing 部份我們從 25 部正當影片取出 46283 筆影片片段與 25 部不當影片取出 35098 筆影片片段給予標記當作 testing data，預測結果如下表所示。

表 3 影片片段偵測預測分類準確率

種類\種類	正當影片片段	不當影片片段
正當影片 片段	98.81% (45730/46283)	1.19% (553/46283)
不當影片 片段	10.25% (4681/45645)	89.75% (40964/45645)

表 4 圖片過濾器分類準確率

種類\種類	正當影片片段	不當影片片段
正當圖片	99.34% (46450/46758)	0.66% (308/46758)
不當圖片	41.13% (13508/32841)	58.87% (19333/32841)

我們觀察表 3 表 4，在正當影片片段與正當圖片分類的準確率個別為 98.81% 與 99.34%，準確率些微下降了 0.53%；而在不當影片片段與不當圖片的準確率分別為 89.75% 與 58.87%，準確率卻提昇了 30.88%。在影片片段的偵測方法我們保持了圖片過濾器過濾正當圖片的偵測高準確率，也改進並提高了圖片過濾器過濾不當圖片的準確率。

5.2 以影片片段色情程度分類影片

在影片偵測的模式中，我們希望正當影片出現誤判的影片片段越小越好。我們比較正當影片與不當影片中有出現最大的影片片段(如圖 8、圖 9 所示)，並觀察在幾個大小的影片片段內分類正當與不當影片的準確率(如圖 10)。

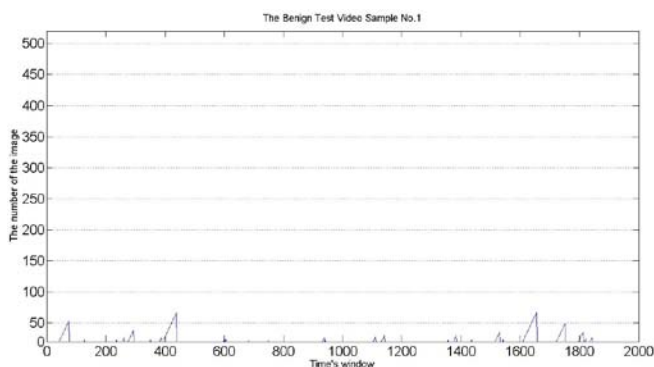


圖 8 正當影片偵測結果示意圖

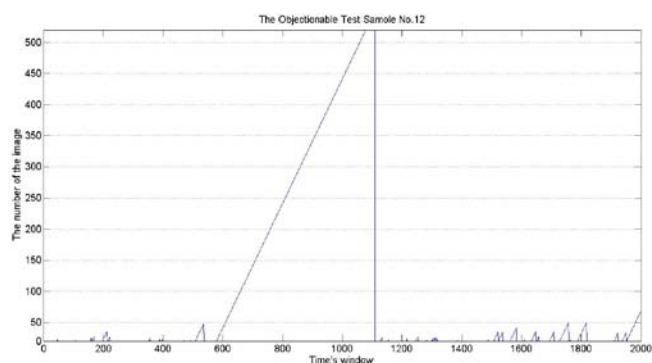


圖 9 不當影片偵測結果示意圖

由圖 8 來說明，正當影片的影片片段偵測，我們發現正當影片片段被誤判為不當影片片段的影片片段大小，較少超過 50 張(2 分 30 秒)圖片，甚至都不超過 100 張(5 分鐘)圖片；而不當影片(如圖 9)會在某一個時間點上使影片片段慢慢增加圖片張數，甚至超過 500 張(25 分鐘)圖片張數，也就是說，此段影片片段連續偵測到 480 次(500-20) 以上的影片片段被分類到不當影片片段。

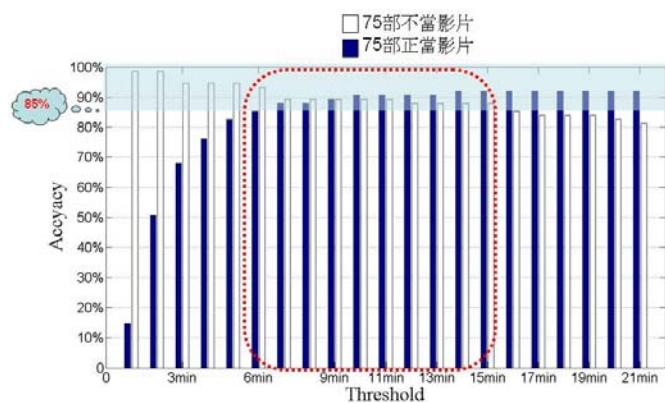


圖 10 不當影片偵測結果示意圖

由圖 10 觀察發現，在正當影片與不當影片兩者分類都有超過 85% 的準確率的門檻在 6 分鐘到 16 分鐘之間，也說明了以 6 分鐘到 16 分鐘的影片片段大小來區分正當與不當影片的各別準確率高於 85%，而最快也能在至少偵測影片 6 分鐘的影片片段就能判別出此影片片段是否為正當影片片段或者為色情影片。

六、結論與未來方向

實驗結果發現，在只用圖片過濾器分類不當影像分類正當影像與不當影像的總準確率為 82.64%，而使用了圖片過濾器與影片片段偵測的方法可將準確率提高到總準確率為 94.31%，提昇了 11.67% 的準確率，說明了使用影片片段偵測法來分類影片屬於正當影片或不當影片的正確率也會有效的提高準確率。

在影片分類方面，利用門檻值來分類正當影片片段與不當影片時，不當程度在 6 分鐘到 16 分鐘的影片片段大小來區分正當與不當影片個別的準確率高於 85%，而最快也能在至少偵測影片 6 分鐘的影片片段就能判別出此影片片段是否為正當影片片段或者為不當影片。

由於本文所設計的影片片段偵測方法，在偵測到不當影片片段中還是會包含某些程度的正當影片資訊，如何降低正當影片資訊出現在不當影片片段中，降低犧牲掉的正當影片資訊，也是未來必須探討的重要議題。

我們將尋找正當影片片段與不當影片片段之間還有哪些特徵上的差異，如聲音信號，或者影像處理上的特徵當作 SVM 的屬性，藉以幫助預測與分類，提昇分類的準確率。

本文探討不當影片的部份，多是以黃種人的膚色資料為主，對於其他膚色類型的不當影片，是無法運用本研究所完成的系統過濾出含有不當資訊的影片，未來也必須結合多個分類器的方式，去針對不同膚色的不當影片做過濾，已達到更完整的過濾不當串流影片。

七、參考文獻

- [1] 林柏璋，使用階層式支撐向量機分類器過濾不當影片。台灣科技大學資訊工程系碩士論文，民 97。
- [2] Yi-Leh Wu, Edward Y, Cheng, Kwang-Ting

Cheng, Cheng-Wie Chang, Chen-Cha Hsu, Wie-Cheng Lai, and Ching-Tung Wu. "MORF: A Distributed Multimodal Information Filtering System," *In Proceedings of the third IEEE Pacific-Rim Conference on Multimedia (PCM 2002)*, Pages 279-286, Taiwan, December 2002.

- [3] Qiang Zhu, Ching-Tung Wu, Kwang-Ting Cheng, Yi-Leh Wu. "An Adaptive Skin Model and Its Application to Objectionable Image Filtering," *In Proceedings of the 12th annual ACM international conference on Multimedia*, Pages: 56-63, 2004.
- [4] R. Collobert and S. Bengio. "Svmtorch: Support Vector Machines for Large-Scale Regression Problems," *Journal of Machine Learning Research*, vol.1, pp.143-160, Spring 2001.
- [5] *Support Vector Machine 簡介* <http://www.cmlab.csie.ntu.edu.tw/~cyy/learning/tutorials/SVM3.pdf>
- [6] Chih-Wei Hsu, Chih-Chung Chang, and Chih-Jen Lin. "A Practical Guide to Support Vector Classification," 2008. <http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>
- [7] Chih-Chung Chang and Chih-Jen Lin. "LIBSVM: A Library for Support Vector Machines," 2001. Software is available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>