

Deep Learning – Assignment I

楊子萱

Institute of Data Science
National Cheng Kung University
Tainan, Taiwan
re6114056@gs.ncku.edu.tw

Abstract—利用 image 資料集進行影像分類的任務，共使用三種不同的特徵抽取方法：Color histogram（顏色直方圖）、Edge Detection（邊緣檢測）和 Hog（方向梯度直方圖）去提取影像的不同特徵，比較不同種的特徵抽取對於預測的效果。

使用多種分類器先對資料集進行增量學習，再將分類效果最好的與集成學習的分類器對整個訓練集進行訓練。發現分類器中，Xgboost 的效果最好，其次是 Random forest，MultinomialNB 的表現也相當不錯。Xgboost 在於 Top1 Accuracy 中最高可達到 0.15 的準確率，Top5 Accuracy 最高可以達到 0.38。

Keywords—Image Classification, feature extraction, Color histogram, MultinomialNB, Random forest, Xgboost

I. INTRODUCTION

利用 image 資料集去從事影像分類的任務，共使用了三種不同的特徵抽取方法：Color histogram（顏色直方圖）、Edge Detection（邊緣檢測）和 Hog（方向梯度直方圖）去提取影像的不同特徵，最終再進行分類。

先透過 Sklearn 增量學習中的三種不同的分類器：SGDClassifier、Perceptron 和 MultinomialNB 進行分類預測。這些分類器可以在不需要重新訓練整個模型的情況下，對新的數據進行學習和分類。經過實驗，我們發現 MultinomialNB 在增量學習中的效果最好。

最後，我們將 MultinomialNB、Random forest 和 Xgboost 針對整個訓練集進行訓練，並將它們應用於測試集進行評估。我們發現，這三種分類器中，Xgboost 的效果最好，其次是 Random forest，MultinomialNB 的表現也相當不錯。

II. FEATURE EXTRACTION

A. Color histogram（顏色直方圖）

Color histogram 是將一張圖片中的像素按照其顏色值進行統計，並以直方圖的形式呈現。並且每個直方圖的 bin 區間表示一個顏色值的範圍，而每個 bin 中的值表示該顏色值的像素數量，若是 RGB 影像則會將每個 Channel 分開統計。

Color histogram 可以用來進行多種影像處理任務，如顏色分割、顏色增強和影像檢索等。其中，顏色分割是指將圖片中的不同顏色區域分開，截取特定 channel 的顏色當作特徵向量，顏色增強則是指通過調整已計算完的直方圖來調整圖片的對比度和色彩平衡，而影像分類則是通過比對兩幅圖片的 color histogram 來評估它們的相似度。

因此在本次影像分類任務中採用 color histogram 方法去收集各個圖片的顏色直方圖，為了整體的處理方便性與一致性，統一將 RGB 圖片轉換成灰階圖片去做後續的特徵萃取，再來因為 bin 數量會影響直方圖的分辨率，bin 數量越多，直方圖的分辨率越高，因此此分類任務中 bin 採用 256，希望能夠將顏色全部的分類出來，以提高分辨正確率。

B. Edge Detection（邊緣檢測）

Edge Detection 用於檢測影像中物件的邊緣或輪廓。邊緣通常是指影像中顏色、亮度或紋理等特徵發生突然變化的位置，邊緣檢測可以提取影像中的重要特徵，並進一步進行影像分割、物體識別等任務。

邊緣檢測常用的算法包括 Sobel、Prewitt、Canny 等。Sobel 和 Prewitt 算法是一種基於梯度的邊緣檢測方法，它們通過計算影像在不同方向上的梯度值來檢測邊緣。Canny 算法則是一種更複雜的邊緣檢測算法，它同樣基於梯度，但通過多步驟的處理來實現更精確的邊緣檢測效果。

因此在本次任務中採用 Canny 算法，希望能夠萃取出較全面且較精確的邊緣特徵以去更好的預測出類別，但邊緣檢測容易受到影像噪聲的影響，而在此資料集中並未先做物件偵測等的預處理，因此在目標影像中包含許多其餘物件，而導致預測準確性下降。

C. Hog（方向梯度直方圖）

HOG 是 Histogram of Oriented Gradients 的縮寫，是一種透過方向梯度來表示圖像特徵的方法，常被用於視覺影像任務中，例如目標物件檢測和人臉識別。這個方法

因具有高魯棒性和計算效率高的特點而在人臉識別、行人檢測等領域廣泛應用。

HOG 算法將圖像分成多個小單元(cell)，計算每個單元內像素的梯度方向(Gradient Direction)和梯度幅度(Gradient Magnitude)，然後根據每個單元內的梯度方向值將梯度幅度劃分到該方向區間中，透過所有單元的梯度方向直方圖去表示整個圖像的特徵。最後，由於 HOG 特徵容易受到光照、噪聲等因素影響，因此會將 K 的小單元組成一個區塊(block) 一起作 Normalization，去提升算法的魯棒性。

在實際使用中，可因應不同的分類任務調整 block size、block stride、cell size 等參數來優化 HOG 的特徵萃取能力。而在本次任務中將全數影像調整為 64 x 64，以方便後續處理與預測，cell size 設為 8 x 8，block size 設為 16 x 16，使得整體的維數不會過大，讓預測階段不會因維度過大而無法運算。

III. CLASSIFIER

A. Incremental Learning

在本次影像辨識任務中採用了增量學習(Incremental Learning)，去比較不同 Training example 對於訓練 Accuracy 的影響，增量學習是一種在已有的模型基礎上不斷地學習新的數據，而不需要重新訓練整個模型的方法，可以快速地將新數據加入到已有的模型中，但不會影響原有的模型性能。此方法可以減少存儲和計算資源的消耗、提高模型的性能和效率、降低過擬合的風險等。

由於透過 Sklearn package 因此選擇了三種在 sklearn 中能夠進行 partial fit 的 classifier 以下逐一介紹：

I. SGDClassifier

一個基於隨機梯度下降 (SGD) 優化算法的線性分類器，每次僅處理一個樣本，因此使得 SGDClassifier 在需要處理大量數據時非常有效率。而且能夠使用不同的 loss function 進行優化，以避免過度擬合的問題，。

II. Perceptron

Perceptron 是一種最簡單的神經網絡模型，接受一系列向量輸入，對每個向量進行加權求和，然後通過一個 activate function 進行分類。如果輸入向量的加權和大於 threshold，則 Perceptron 將其分類為 positive，否則為 negative。

雖然 Perceptron 為簡單的模型，當問題數據集為線性可分時分類準確性會很高，而為非線性可分時，可能會造成無法收斂的情況。

III. MultinomialNB

是一種基於貝式定理的分類算法，此模型學習每個類別的特徵分佈，並使用貝式定理進行分類，優點為計算效率高，適用大型數據集，且不需要大量樣本數據就可以得到合理的分類結果，因此在樣本數據較少的情況下也可以使用，此結果在後續的 training example 的圖表中也可以看見。

B. Ensemble Learning

本次採用了兩個常見的集成學習分類器，Random Forest 以及 Xgboost，Ensemble Learning 是利用多個分類器預測結果集成來提高模型的性能，減低數據偏差以及噪聲造成的影響。

I. Random Forest

Random Forest(隨機樹森林) 是一種基於 Bagging 的集成方法。結合多顆決策樹做預測，最後將各顆樹預測的結果進行平均或投票來獲得最終的預測結果。因為具有結合多顆決策樹的性質，因此在於缺失以及噪聲的處理上表現出色，提高模型的穩定性，並且因為決策樹能夠計算各個特徵的重要性排名，可以易於了解特徵對於分類時的效果。

II. XGBoost

是以 Gradient Boosting 為基礎，結合 Bagging 和 Boosting 的優點。使得每一棵樹是相互關聯，而後面生成的樹能夠修正前面一棵樹犯錯的地方，並且採用特徵隨機採樣的方式進行，因此在準確率的表現上很好，且因為透過梯度進行計算與優化，因此在於運算上也非常快速。

IV. EXPERIMENT

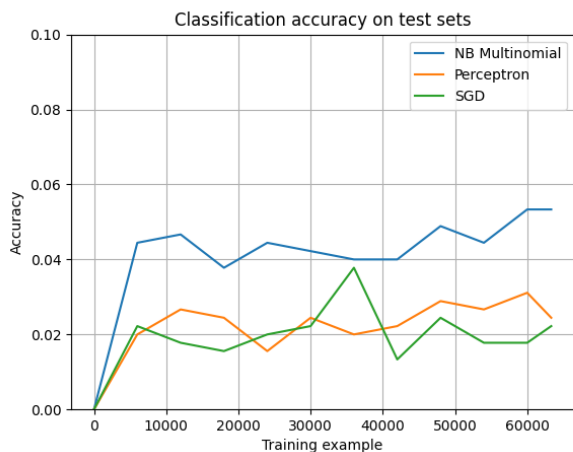
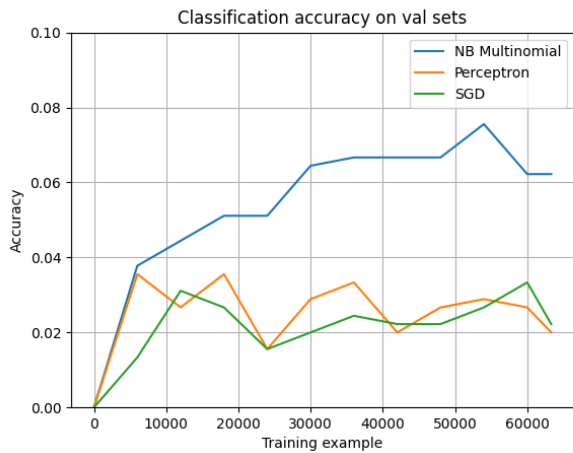
在本次試驗中，先進行增量學習，並且在增量學習的三種分類器中選擇效能最好的，與後續的強分類器 Random Forest 以及 XGBoost 去做第二階段的比較。

I. Effect of Training example on different classifier

在增量學習的分類器中 MultinomialNB 在 Training example 為 54000 張 image 時 validation set 可以達到 0.076 的準確率，而 testing set 中可達到 0.044 的準確性，而當 Training example 為 63325 張全部的 image 時，testing set 可提升到 0.053 的準確性，validation set 則降至 0.062 的準確率，顯示再多加進來 training example 後降低了過擬合的效果。

下圖中，MultinomialNB 比起其餘兩種的分類效果來得好，且在於使用少量數據集進行預測時，準確率能夠

較快速的上升，可以看出當 training example 為 6000 時，預測的準確性就已經高於 SGD 以及 Perceptron 的分類效果了，因此選用 MultinomialNB 與其餘較強的集成學習分類器一起針對全部的 Training data 做模型的訓練與預測，會呈現在下一個部分。



II. Effect of Different Feature Extraction on Accuracy

而在第二階段的模型訓練與預測中，將三種特徵萃取器，分別是 Color histogram（顏色直方圖）、Edge Detection（邊緣檢）和 Hog（方向梯度直方圖）分別獨立使用以及進行各種組合，共計得到 7 種不同的組合方式，最終再透過三個較強的分類器 MultinomialNB、Random Forest、XgBoost 對模型做訓練，並檢測正確性。可以在下頁的表格中看見完整的實驗結果，CH 表示為 Color histogram，ED 表示為 Edge Detection。

對於 MultinomialNB 分類器而言，HOG 是非常好的特徵萃取方式，比起搭配其他的特徵萃取方法(CH、ED)，單一種在於不管是 validation set 還是 testing set 皆達到最高的準確率，Top-1 分別是 0.067 與 0.078，而對於 Random Forest 而言，在於 validation set 上效果最好的為 Color histogram，而三種的混和則在於 testing set 上效果最好，但也在 validation set 擁有蠻高的準確率，最後，正確性最高的為 XgBoost，將三種特徵萃取方式混合後，在於 Top-1 Accuracy 中可以達到 0.15 的準確性，而在

Top-5 Accuracy 更可以達到 0.38 的準確性，顯示基本上在於預測類別中機率最高的 5 個類別中屬於正確類別的機率有 4 成。

III. Runtime

統計每一個特徵萃取方式在於不同分類器訓練模型時的 Runtime，可於下頁中的表格內查看。

可以明顯的看出 MultinomialNB 的訓練時間非常的短，雖然分類的正確性不如其餘兩種，但也可以達到一定的效果，且在第一階段中發現，MultinomialNB 不需要使用大量的資料集即可達到不錯的分類效果，因此若是在於小量的資料集或是對於訓練時間的需求較高時，只要找到適當的特徵萃取方式，也 MultinomialNB 能夠是很好的分類器。

訓練時間最長的為 XgBoost，其次為 Random Forest，因為兩種皆為集成學習，會結合多顆決策樹做最終決策，因此訓練時間會比起一般分類時間來得久。

V. CONCLUSION

在於本次實作結果上，使用了多種不同的分類器與特徵萃取方式，希望能夠從中去了解不同特徵萃取方式對於影像特徵向量建構上的影響，並且也了解不同的分類器較適合哪種的特徵萃取方法。

期望未來能夠再結合多中不同的特徵萃取方法，去找到能夠更加找出影像特徵的萃取方式，或者多嘗試不同的模型或者嘗試不同的參數去優化整體的預測效率。

本次實作遇到的困難：

1. 因為影像大小各不同，因此若是直接使用影像去取特徵時，會造成特徵向量長度不一的問題，而無法做分類預測，但若是直接調整大小也不太知道取何種大小能夠最符合全部的影像，且不會使得太多的資訊流失。
2. 許多影像中不是僅包含目標物件，因此在於特徵萃取時，會含有許多不需要的資訊，導致最終的預測準確性下降，且每一個類別資料夾中可能含有不是該類別的影像，標記的品質也可能影響預測的準確性。
3. 由於資料集含有 6 萬多張影像，因此若是使用訓練量較龐大的分類器會導致運算資源不足，因此在分類器的選擇上也需要再多加考量。

GitHub 網址：

https://github.com/tenyang1999/DL_Assignment_1-Image_Classification

實驗結果

Validation set

	MultinomialNB		XgBoost		Random Forest	
	Top1	Top5	Top1	Top5	Top1	Top5
CH	0.0622	0.1289	0.1089	0.3111	0.1133	0.3156
ED	0.0489	0.1711	0.0422	0.1822	0.0356	0.1267
Hog	0.0667	0.2667	0.0844	0.3222	0.0711	0.2467
CH+ ED	0.0622	0.1356	0.1044	0.3356	0.0956	0.2978
CH+ Hog	0.0622	0.1289	0.1378	0.4111	0.1089	0.2978
ED+ Hog	0.0511	0.1756	0.0911	0.3089	0.0822	0.28
CH+ED+ Hog	0.0622	0.1356	0.14	0.3844	0.1	0.3022

Testing set

	MultinomialNB		XgBoost		Random Forest	
	Top1	Top5	Top1	Top5	Top1	Top5
CH	0.0467	0.1356	0.0733	0.2467	0.0756	0.24
ED	0.0378	0.16	0.04	0.1667	0.0289	0.1356
Hog	0.0778	0.2644	0.1156	0.3533	0.08	0.2867
CH+ ED	0.0533	0.1311	0.0756	0.2533	0.0711	0.2267
CH+ Hog	0.0467	0.1356	0.1222	0.3733	0.0978	0.2867
ED+ Hog	0.04	0.1622	0.1267	0.3333	0.0711	0.2333
CH+ED+ Hog	0.0533	0.1311	0.1467	0.3867	0.1089	0.3089

Runtime

	MultinomialNB	XgBoost	Random Forest
CH	0.09	102.1s	153.53s
ED	0.11s	87.33s	18.45s
Hog	0.14s	525.79s	435.37s
CH+ ED	0.13s	179.01s	116.12s
CH+ Hog	0.15s	610.02s	416.41s
ED+ Hog	0.18s	1026.13s	365.55s
CH +ED+ Hog	0.31s	624.61s	363.53s