

Learning Gender with Boosting

指導教授：楊致芳

學生：謝文哲

摘要

- 討論人臉性別樣本的特性如何影響LogitBoost與AdaBoost，評估兩種機器學習的演算法在性別辨識問題上的表現
- 人臉性別辨識系統實作
- 關鍵字：Face Detection、Gender Classification、Feature Extraction、Boosting、Machine Learning

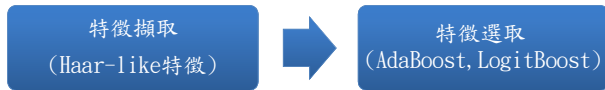
什麼是人臉性別辨識？

- 人臉辨識是讓電腦在一張影像中判斷人臉可能的位置和大小，忽略其它不是人臉的東西像是身體、建築物等
- 性別辨識是給定一張人臉，透過過去的經驗計算藉以判斷人臉是男性還是女性

實驗方法

性別辨識的問題解決策略-物件檢測架構

- 特徵擷取是用特徵從影像中擷取數據
- 特徵選取則是利用機率統計的方法從數據中自動分析以獲得規律，找出對檢測有幫助的數據

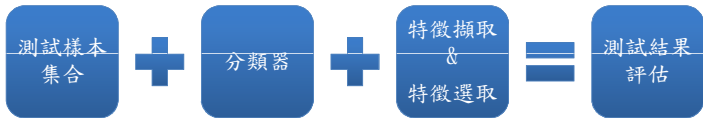


測試程序和訓練程序

- 解決性別辨識問題需要兩種類別的樣本：男性和女性
- 將大部分的樣本做為訓練樣本集合，經過特徵擷取和特徵選取的訓練程序可獲得分類器，分類器蒐集了對性別辨識有幫助的資訊



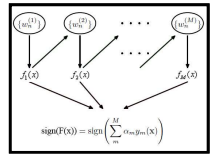
- 將剩下的樣本做為測試樣本集合，透過測試程序可以評斷分類器的好壞



文獻探討

Boosting

- 監督式學習法
- 循序挑選弱分類器組成強分類器，弱分類器的預測能力須大於隨機猜測
- 由弱分類器的表現決定樣本權重的調整方式



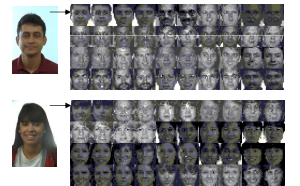
AdaBoost vs. LogitBoost

- AdaBoost minimize $E(e^{-yF(x)})$
- LogitBoost minimize $E \log(1 + e^{-2yF(x)})$
- AdaBoost is sensitive to noisy data and outliers.

實驗與討論

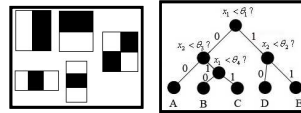
Color FERET人臉資料庫

- 只取正面人臉
- 利用人臉辨識擷取人臉樣本，將樣本縮放至20x20的大小

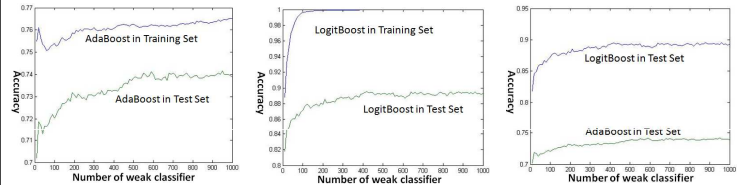


實驗參數

- Color FERET資料庫，樣本大小：20x20像素
- 使用Haar-like特徵和決策樹建構弱分類器
- 將男性與女性樣本的比例調至1比1
- Cross Validation
- 透過分類器在訓練樣本集合的表現評估實驗的正確性，藉由分類器在測試樣本集合的表現判斷哪種Boosting所訓練的分類器適合解決性別辨識問題
- LogitBoost性別分類器獲得了89.31%的正確率，高於AdaBoost性別分類器的73.89%
- 由C語言撰寫實驗系統



分類器	辨識正確率
	測試樣本集合
AdaBoost	73.89%
LogitBoost	89.31%



建構一個人臉性別辨識系統

系統所使用的技術

- C++、OpenCV 2.2、Wxwidgets 2.9.2
- CMake 2.8、GCC 4.5.2
- 跨平台就緒

如何改善OpenCV人臉辨識系統

- 適當的旋轉影像並蒐集偵測結果，刪除重複偵測的區域
- 用人臉膚色刪除辨識錯誤的區域，將影像色彩空間從RGB轉換至YCrCb，統計膚色像素所占比例，淘汰比例過低的區域

在實作上導致性別辨識錯誤的因素

- 人臉辨識的步驟可能抓取一些側臉或是傾斜角度過大的人臉，在性別辨識之前必須對擷取的人臉進行前處理
- 現實生活中的影像亮度不一，不像人臉資料庫在拍攝的過程保持一定的亮度

