

# 應用監督式分類於多時期嘉市綠地面積比例 分析比較

R05521121 陳立恆

## 一、前言及動機

身為嘉義長大的嘉義小孩，每天上下學除了觀察路上形形色色的人外，最常被嘉義人閒話家常的是，某某地方又開始要蓋公園了。依內政部數據指出 106 年 1 月止全國 65 歲以上老年人口比已高達 13.27%，嘉義市更攀升到 13.52%，其中東區老人人口比 14.35%，已正式達到「高齡社會」標準(14%)。因此在高齡人口這麼高的地區，高齡友善健康城市就是嘉義市政府追求的目標之一。

政府表示為了增加老年人口的活動空間，以及為了健康目的，公園綠地讓老人泡茶聊天、健走散步的地方是不可或缺的。對比下圖，六都公園綠地面積及每人平均值，嘉市每人平均值僅不到 2 平方公尺，遠低於六都的每人平均值。因此在 95 年黃敏惠市長上任後，共開闢 11 處都市計畫公園，包括文化公園、湖內兒童公園、短竹公園、劉厝公園、劉厝 2 號、劉厝 3 號公園及嘉雄陸橋下番仔溝公園（公 13、14）、中庄公園、北社尾公園及光路公園，公園面積共計增加 11.94 公頃，每位市民平均享有的公園面積因此增加 0.44 平方公尺，由每人 1.41 平方公尺，增加至 1.85 平方公尺，每人享有公園綠地面積增加率為 31.2%，期許能改善此一狀況。以及 2005 年相比其他縣市平均每人綠地量也是非常稀少的

六都都市計畫區 公園綠地面積及每人平均值		
排序	面積 (公頃)	每人平均值 (平方公尺)
高雄市	3188.84	11.47
台中市	2028	7.45
台南市	1574.97	8.35
新北市	1345.76	3.39
台北市	1309.65	4.85
桃園市	642.52	3.11

圖一、2008 年六都公園面積

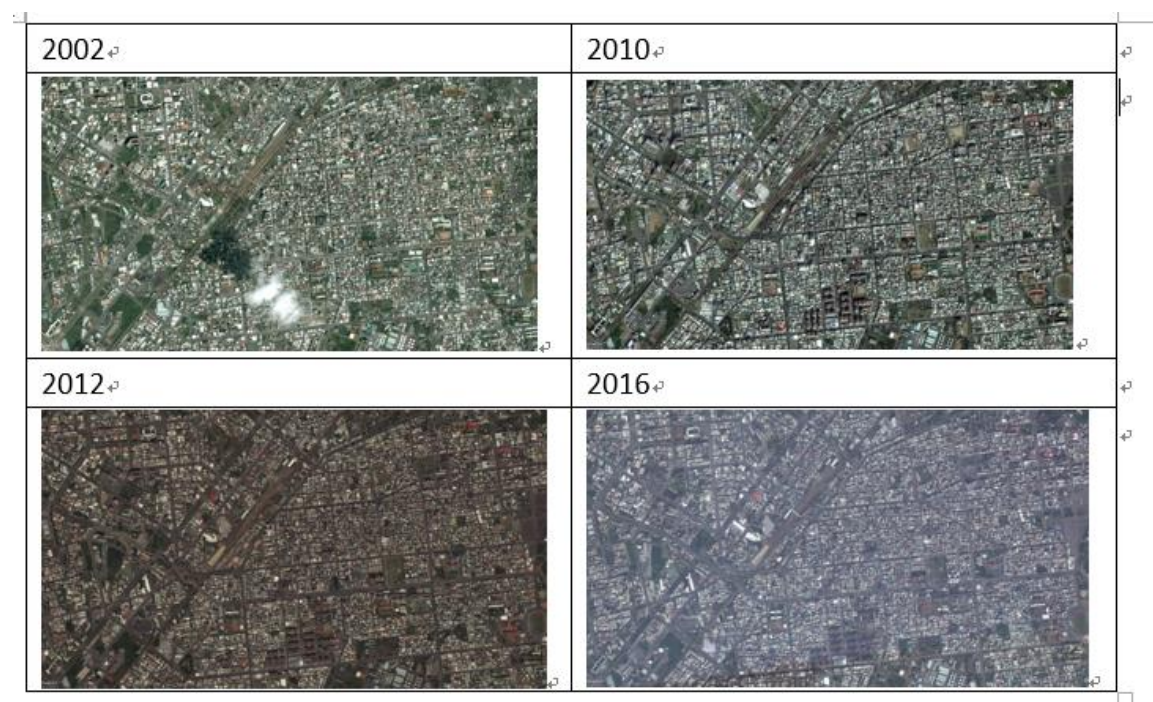
地區別	公園 (公頃)	綠地 (公頃)	廣場 (公頃)	兒童 遊樂場 (公頃)	總計(公頃)	人口數	平均每人綠地量 (平方公尺/人)
基隆市	259.60	21.63	0.16	3.08	284.47	391,727	7.26
新竹市	214.02	44.97	5.23	4.74	268.96	390,692	6.88
臺中市	1,100.97	41.51	3.91	61.74	1,208.13	1,032,778	11.69
嘉義市	107.80	28.96	1.44	6.23	144.43	271,701	5.31
臺南市	688.78	267.35	17.01	25.03	998.17	756,859	13.18

圖二、2005 年嘉義市平均每人綠地量

因此預計利用遙測影像的監督式分類，觀測嘉市公園綠地面積的變化，希望能以遙感探測的方式，應證上述公園綠地的面積的增加，或是即便政府蓋了許多公園，綠地面積還是沒有增加，是否政府蓋公園有其他目的。並搭配嘉義市人口，取得每人平均值，是否如上述數據。或者由於台灣的都市化日趨劇烈，因此嘉市由於人口嚴重外移，產生許多幽靈人口，導致每人平均綠地面積遠被低估，亦是探討的課題之一。

## 二、場景選擇

由於嘉義市的衛星影像來源非常稀少，較難找到含有 4 個波段(R、G、B、IR)的各個時期的嘉義市衛星影像。因此影像來源為 Google Earth 上，2002 年、2010 年、2012 年和 2016 年的嘉義市區衛星影像，波段數為 RGB 三個波段。由於來源為 Google Earth，其影像有一些雲層遮蔽，以及影像鑲嵌時調色不均的問題，如圖三所示，而這些因素將會對後續的成果造成影像。



圖三、不同時期的嘉義市衛星影像

Image Height	4800
Image Width	2649
GSD	0.75m
Bands Number	3(R、G、B)

表一、衛星影像資訊





### 三、方法介紹



圖四、研究流程圖

#### I. 訓練區選擇

由於本次 project 預計利用監督式學習，進行影像的分類，監督式學習是一個機器學習中的方法，可以由訓練資料中學到或建立一個模式(learning model)，並依此模式推測新的資料。訓練資料是由輸入和預期輸出所組成。函數的輸出可以是一個連續的值(也就是迴歸分析 regression)，或是預測一個分類標籤(也就是分類, classification)。一個監督式學習者的任務在觀察完一些訓練資料後，去預測這個函數對任何可能出現的輸入的值的輸出。要達到此目的，學習者必須以"合理"的方式從現有的資料中一般化到非觀察到的情況。在人類和動物感知中，則通常被稱為概念學習。因此訓練區的選擇對於監督式學習非常重要。

2002	2010
	
2012	2016
	



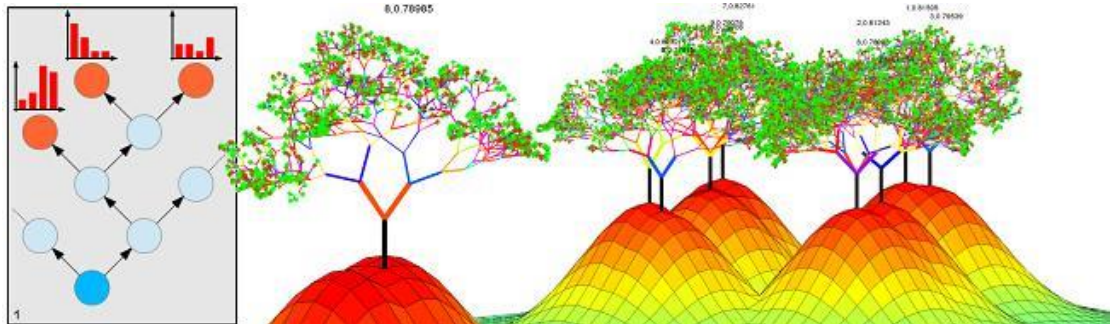
表二、各年度三個訓練區選擇

## II. 訓練方式➡機器學習

分為隨機森林以及最大似然法兩部分。

### • Random Forest

隨機森林（random forest）顧名思義，是用隨機的方式建立一個森林，森林裡面有很多的決策樹組成，隨機森林的每一棵決策樹之間是沒有關聯的。在得到森林之後，當有一個新的輸入樣本進入的時候，就讓森林中的每一棵決策樹分別進行一下判斷，看看這個樣本應該屬於哪一類，然後看看哪一類被選擇最多，就預測這個樣本為那一類。




圖四、隨機森林示意圖

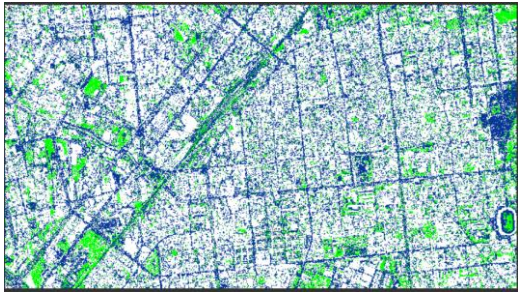

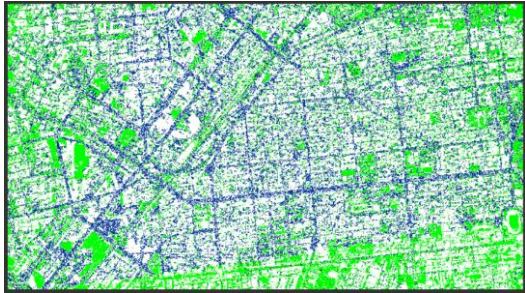
### • Maximum Likelihood

是遙測中較傳統的監督式分類方法，最大似然法為利用統計的方式。其假設所有類別資料皆具有高斯常態分佈特性，因此訓練樣本的統計值即為光譜平均值及其協變方矩陣，而辨識函數則是某一像元屬於某一類別的似然度，根據似然度大小即可進行分類。



## 四、成果分析

### ➤ Random Forest

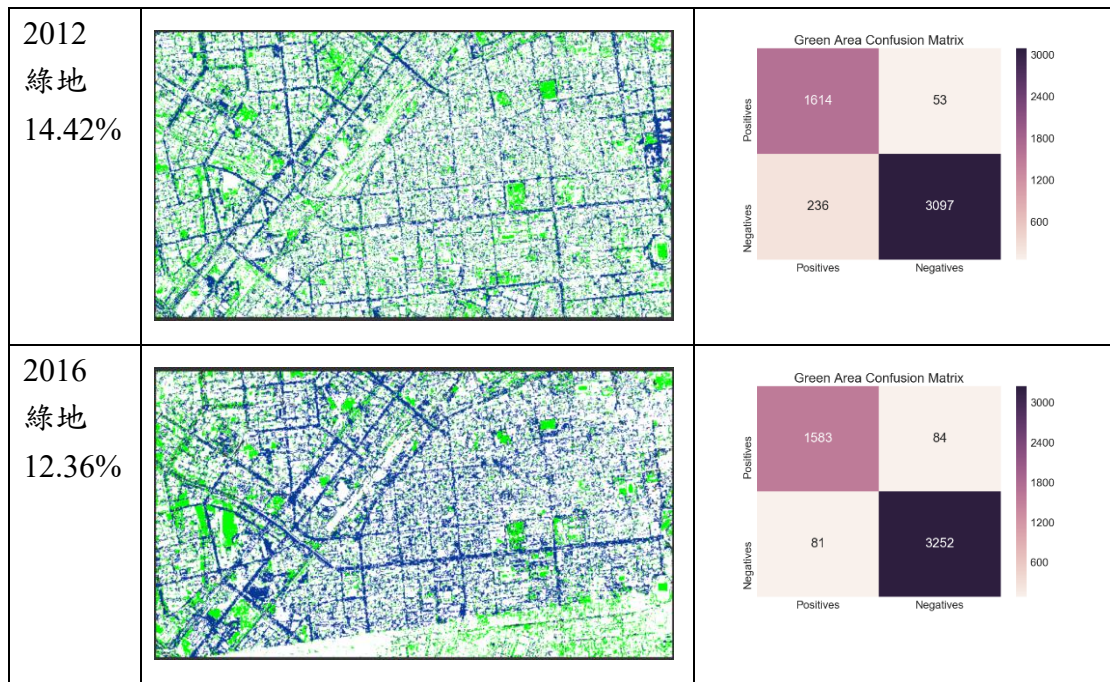
年份	分類成果	綠地分類混淆矩陣									
2002 綠地 26.68%		<p>Green Area Confusion Matrix</p> <table border="1"> <tr> <td>Positives</td><td>1553</td><td>114</td></tr> <tr> <td>Negatives</td><td>121</td><td>3212</td></tr> <tr> <td></td><td>Positives</td><td>Negatives</td></tr> </table> <p>Color scale: 600 to 3000</p>	Positives	1553	114	Negatives	121	3212		Positives	Negatives
Positives	1553	114									
Negatives	121	3212									
	Positives	Negatives									

年份	分類成果	綠地分類混淆矩陣									
2010 綠地 10.32%		<p>Green Area Confusion Matrix</p> <table border="1"> <tr> <td>Positives</td><td>1528</td><td>138</td></tr> <tr> <td>Negatives</td><td>165</td><td>3169</td></tr> <tr> <td></td><td>Positives</td><td>Negatives</td></tr> </table>	Positives	1528	138	Negatives	165	3169		Positives	Negatives
Positives	1528	138									
Negatives	165	3169									
	Positives	Negatives									
2012 綠地 13.73%		<p>Green Area Confusion Matrix</p> <table border="1"> <tr> <td>Positives</td><td>1517</td><td>150</td></tr> <tr> <td>Negatives</td><td>158</td><td>3175</td></tr> <tr> <td></td><td>Positives</td><td>Negatives</td></tr> </table>	Positives	1517	150	Negatives	158	3175		Positives	Negatives
Positives	1517	150									
Negatives	158	3175									
	Positives	Negatives									
2016 綠地 22.27%		<p>Green Area Confusion Matrix</p> <table border="1"> <tr> <td>Positives</td><td>1612</td><td>55</td></tr> <tr> <td>Negatives</td><td>59</td><td>3274</td></tr> <tr> <td></td><td>Positives</td><td>Negatives</td></tr> </table>	Positives	1612	55	Negatives	59	3274		Positives	Negatives
Positives	1612	55									
Negatives	59	3274									
	Positives	Negatives									

➤ **Maximum Likelihood**

年份	分類成果	綠地分類混淆矩陣									
2002 綠地 21.39%		<p>Green Area Confusion Matrix</p> <table border="1"> <tr> <td>Positives</td><td>1597</td><td>70</td></tr> <tr> <td>Negatives</td><td>257</td><td>3076</td></tr> <tr> <td></td><td>Positives</td><td>Negatives</td></tr> </table>	Positives	1597	70	Negatives	257	3076		Positives	Negatives
Positives	1597	70									
Negatives	257	3076									
	Positives	Negatives									
2010 綠地 12.20%		<p>Green Area Confusion Matrix</p> <table border="1"> <tr> <td>Positives</td><td>1587</td><td>79</td></tr> <tr> <td>Negatives</td><td>145</td><td>3189</td></tr> <tr> <td></td><td>Positives</td><td>Negatives</td></tr> </table>	Positives	1587	79	Negatives	145	3189		Positives	Negatives
Positives	1587	79									
Negatives	145	3189									
	Positives	Negatives									

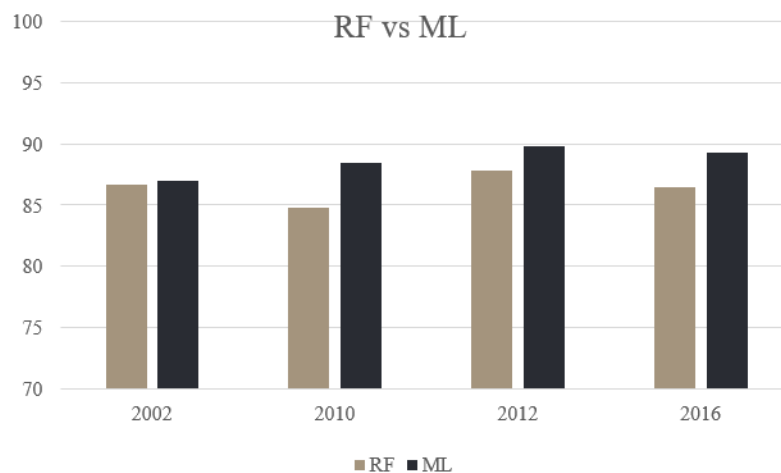




➤ **Using Testing dataset, correct predictions comparison**

(%)	Random Forest	Maximum Likelihood
2002	86.00	86.96
2010	84.84	88.42
2012	87.82	89.80
2016	86.44	89.24

表三、隨機森林及最大似然法正確預測比率



圖五、隨機森林及最大似然法正確預測比率

➤ 嘉義市綠地面積、每人平均綠地面積

Area	R F (公頃)	ML (公頃)	m <sup>2</sup> /人	R F	ML
2002	200.07	160.38	2002	7.41m <sup>2</sup>	5.94m <sup>2</sup>
2010	77.49	91.53	2010	2.87m <sup>2</sup>	3.39m <sup>2</sup>
2012	102.81	108.27	2012	3.81m <sup>2</sup>	4.01m <sup>2</sup>
2016	170.37	92.61	2016	6.31m <sup>2</sup>	3.43m <sup>2</sup>

## 五、結論及探討

➤ Random Forest vs Maximum Likelihood

從隨機森林的成果圖，以肉眼直接觀看，發現在 2002 以及 2016 有較大的錯誤；反之在最大似然法，以肉眼觀看可以發現並沒有什麼大錯誤的情形發生。以及從用測試資料隨機森林以及最大似然法的預測精度比較，可以看到在每一年最大似然法的成果都較隨機森林稍稍好一些。但大部分文獻指出最大似然法通常有較差的成果，這次隨機森林成果較差，可能是由於只有影像只有三個波段並且只分為三個訓練區，導致較簡略的最大似然法有較好的精度，而隨機森林或許是模型過於複雜反而有過度擬合的現象。

➤ 嘉義市綠地面積變遷分析

從上面的結果除去紅字錯誤的部分，可看到由 2002 年每人綠地面積大約為 6 平方公尺，隨著空地都拿來蓋建物，每人綠地面積縮減為大約 3 平方公尺，雖然蓋了許多公園，但綠地面積仍然沒增加，主要是因為蓋公園的地方原本就是綠地或空地，因此每人綠地面積不會因為蓋公園而增加。而綠地對於都市來說好比是人類的肺部，是供給城市呼吸不可或缺的地區，故我認為即便一直蓋公園綠地面積沒增加反而還一直縮減，是有些本末倒置。

➤ 總結

基本上 RGB 衛星影像尚可用來做粗略的分析，如果要較精確分析可能需要再加上其他波段如近紅外光或是別的演算法如 SVM 或是類神經網路等等，甚至是搭配地真資料更精準的訓練區圈選才能更正確的進行分類。

## 六、參考文獻

- Camps-Valls, G. (2009). Machine learning in remote sensing data processing. In *Machine Learning for Signal Processing, 2009. MLSP 2009. IEEE International*
- Kulkarni, A. D., & Lowe, B. (2016). Random forest algorithm for land cover classification.

Nitze, I., Schulthess, U., & Asche, H. (2012). Comparison of machine learning algorithms random forest, artificial neural network and support vector machine to maximum likelihood for supervised crop type classification. *Proc. of the 4th GEOBIA*, 7-9.