



PRÁCTICA 3

SIG Y Teledetección

Francisco Martinez Esteso

Junio de 2021

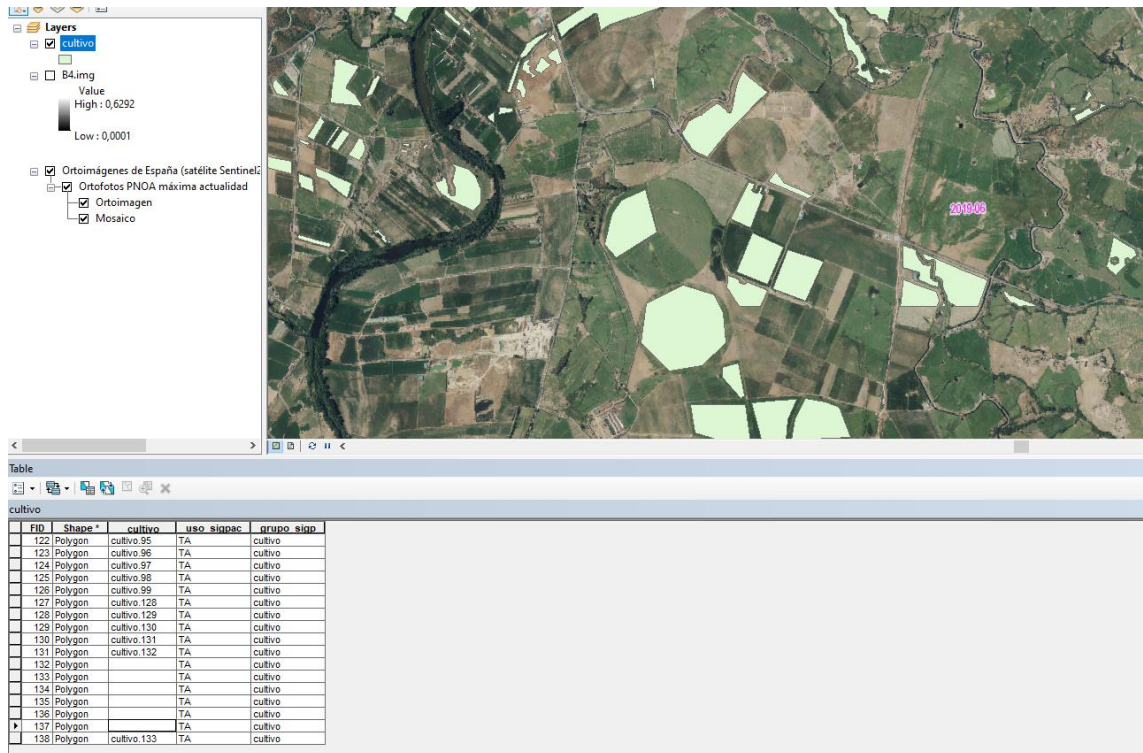
Contenido

Experimentos previos.....	2
Creación de polígonos adicionales.....	2
Variación del número de ejemplos de entrenamiento	3
Variación del número de árboles	3
Variación aleatoria de los parámetros	4
Elección de bandas.....	4
Clasificación óptima	5

Experimentos previos

Creación de polígonos adicionales

Se decide empezar a mejorar el porcentaje de predicciones correctas a partir de la creación de nuevos ejemplos de entrenamiento para el algoritmo, creando nuevos polígonos en las clases.



El resultado es una disminución del porcentaje de acierto en todos y cada uno de los experimentos realizados, siempre con los mismos parámetros para comprobar la validez de los cambios, con los siguientes resultados:

Número de ejemplos introducidos	Árboles	Ejemplos de entrenamiento	% Acierto
5 por clase	10	100	73.9583
7 por clase	10	100	71.6528
10 por clase	10	100	75.7082

```
TruePositives = 8.0000 FalsePositives = 2.0000 TrueNegatives = 82.0000 FalseNegatives = 0.0000
class 7.0: cultivo
accuracy = 0.9375 precision = 0.8000 correlation = 0.7132 errorRate = 0.0625
TruePositives = 8.0000 FalsePositives = 2.0000 TrueNegatives = 82.0000 FalseNegatives = 0.0000

Using Testing dataset, % correct predictions = 73.9583
Total samples = 200
RMSE = 1.9094065395649333
Bias = -0.7291666666666665

Distribution:
class 0.0: forestal      25 (12.5000%)
class 1.0: frutales      25 (12.5000%)
```

Viendo la nula mejora del porcentaje de acierto, la cual se atribuye al desconocimiento en campo de la veracidad de los datos y el poco conocimiento agrícola al vuelo del usuario, se decide continuar con los polígonos originales suministrados por la profesora.

Variación del número de ejemplos de entrenamiento

Se varía el número de ejemplos de entrenamiento en busca de un óptimo para mejorar el porcentaje de acierto. Se prueba con las siguientes configuraciones:

Árboles	Ejemplos de entrenamiento	% Acierto
10	100	84.3750
10	500	52.0879
10	1000	41.8935
10	2500	47.4567
10	5000	37.1429

```

25 TruePositives = 111.0000 FalsePositives = 10.0000 TrueNegatives = 710.0000 FalseNegatives = 14.0000
26 class 7.0: urbano
27 accuracy = 0.9858 precision = 0.9520 correlation = 0.9445 errorRate = 0.0142
28 TruePositives = 119.0000 FalsePositives = 6.0000 TrueNegatives = 714.0000 FalseNegatives = 6.0000
29
30 Using Testing dataset, % correct predictions = 87.5740
31 Total samples = 1690
32 RMSE = 1.082950167407881
33 Bias = -0.10769230769230731
34
35 Distribution:
36 class 0.0: agua 250 (14.7929%)
37 class 1.0: cultivo 250 (14.7929%)
38 class 2.0: forestal 250 (14.7929%)
39 class 3.0: frutales 42 (2.4852%)

```

Variación del número de árboles

Se varía también el número de árboles a crear para buscar de nuevo un óptimo en la clasificación. Se estudian los siguientes casos:

Árboles	Ejemplos de entrenamiento	% Acierto
10	100	84.3750
15	100	89.5833
20	100	82.2917
25	100	83.3333

```

24 accuracy = 0.9583 precision = 0.9000 correlation = 0.8055 errorRate = 0.0417
25 TruePositives = 9.0000 FalsePositives = 1.0000 TrueNegatives = 83.0000 FalseNegatives = 3.0000
26 class 7.0: urbano
27 accuracy = 0.9688 precision = 1.0000 correlation = 0.8510 errorRate = 0.0313
28 TruePositives = 9.0000 FalsePositives = 0.0000 TrueNegatives = 84.0000 FalseNegatives = 3.0000
29
30 Using Testing dataset, % correct predictions = 83.3333
31 Total samples = 200
32 RMSE = 1.1636866703140785
33 Bias = -0.02083333333333348
34
35 Distribution:
36 class 0.0: agua 25 (12.5000%)
37 class 1.0: cultivo 25 (12.5000%)
38 class 2.0: forestal 25 (12.5000%)

```

Variación aleatoria de los parámetros

Se busca ahora crear unos casos de variación de los parámetros para los que obtener un máximo de porcentaje de acierto. Se estudian los siguientes casos:

Árboles	Ejemplos de entrenamiento	% Acierto
15	150	88.6667
25	300	85.7143
13	200	82.1429

```

5 TruePositives = 18.0000 FalsePositives = 6.0000 TrueNegatives = 125.0000
6 class 7.0: urbano
7 accuracy = 0.9667 precision = 1.0000 correlation = 0.8425 errorRate = 0.0
8 TruePositives = 14.0000 FalsePositives = 0.0000 TrueNegatives = 131.0000
9
10 Using Testing dataset, % correct predictions = 88.6667
11 Total samples = 300
12 RMSE = 0.8755950357709131
13 Bias = 0.03333333333333215
14
15 Distribution:
16 class 0.0: agua 38 (12.6667%)

```

Elección de bandas

Se decide eliminar ciertas bandas que no aportan demasiada información en las validaciones cruzadas vistas hasta ahora. Se usan nuestros óptimos en los parámetros de la configuración. Se contemplan los siguientes casos:

Bandas eliminadas	Árboles	Ejemplos de entrenamiento	% Acierto
NDVI45	15	150	81.6597
BI	15	150	65.4629
S2REP	15	150	84.8712

Clasificación óptima

Tras el aprendizaje que nos da nuestros experimentos anteriores ejecutamos un experimento que nos arroja el resultado óptimo con las siguientes configuraciones:

Algoritmo utilizado	Número de ejemplos de entrenamiento añadidos	Bandas eliminadas	Árboles	Ejemplos de entrenamiento	% Acierto
RF	5 a cada clase		15	150	95.7447

```
15 accuracy = 0.9787 precision = 0.9355 correlation = 0.9386 errorRate = 0.0213
16 TruePositives = 29.0000 FalsePositives = 2.0000 TrueNegatives = 109.0000 FalseNegatives = 1.0000
17 class 4.0: urbano
18 accuracy = 0.9929 precision = 0.9677 correlation = 0.9793 errorRate = 0.0071
19 TruePositives = 30.0000 FalsePositives = 1.0000 TrueNegatives = 110.0000 FalseNegatives = 0.0000
20
21 Using Testing dataset, % correct predictions = 95.7447
22 Total samples = 282
23 RMSE = 0.42107596053325946
24 Bias = 0.03546099290780136
25
26 Distribution:
```