

Herramientas de Teledetección Cuantitativa

¿Y de precisión como andamos?

Francisco Nemiña

Unidad de Educación y Formación Masiva
Comisión Nacional de Actividades Espaciales

21 de octubre de 2016



Esquema de presentación

Escenas del capítulo anterior

Filtrado

Técnicas pos-clasificación

Estimación de la precisión

Matriz de confusión

Estimación de áreas

Muestreo

Práctica



La vez pasada vimos

- ▶ Que a partir de esto podíamos definir la ρ_λ la firma espectral como una característica de cada cuerpo.
- ▶ La necesidad de definir categorías de uso y cobertura de forma concisa.
- ▶ La diferencia entre el concepto de categorías de uso y cobertura y clases espectrales.
- ▶ La importancia del espacio espectral para comprender los métodos de clasificación.
- ▶ El funcionamiento del algoritmo k-means de segmentación.
- ▶ El funcionamiento del algoritmo de máxima verosimilitud.



La vez pasada vimos

- ▶ Que a partir de esto podíamos definir la ρ_λ la firma espectral como una característica de cada cuerpo.
- ▶ La necesidad de definir categorías de uso y cobertura de forma concisa.
- ▶ La diferencia entre el concepto de categorías de uso y cobertura y clases espectrales.
- ▶ La importancia del espacio espectral para comprender los métodos de clasificación.
- ▶ El funcionamiento del algoritmo k-means de segmentación.
- ▶ El funcionamiento del algoritmo de máxima verosimilitud.



La vez pasada vimos

- ▶ Que a partir de esto podíamos definir la ρ_λ la firma espectral como una característica de cada cuerpo.
- ▶ La necesidad de definir categorías de uso y cobertura de forma concisa.
- ▶ La diferencia entre el concepto de categorías de uso y cobertura y clases espectrales.
- ▶ La importante del espacio espectral para comprender los métodos de clasificación.
- ▶ El funcionamiento del algoritmo k-means de segmentación.
- ▶ El funcionamiento del algoritmo de máxima verosimilitud.



La vez pasada vimos

- ▶ Que a partir de esto podíamos definir la ρ_λ la firma espectral como una característica de cada cuerpo.
- ▶ La necesidad de definir categorías de uso y cobertura de forma concisa.
- ▶ La diferencia entre el concepto de categorías de uso y cobertura y clases espectrales.
- ▶ La importancia del espacio espectral para comprender los métodos de clasificación.
- ▶ El funcionamiento del algoritmo k-means de segmentación.
- ▶ El funcionamiento del algoritmo de máxima verosimilitud.



La vez pasada vimos

- ▶ Que a partir de esto podíamos definir la ρ_λ la firma espectral como una característica de cada cuerpo.
- ▶ La necesidad de definir categorías de uso y cobertura de forma concisa.
- ▶ La diferencia entre el concepto de categorías de uso y cobertura y clases espectrales.
- ▶ La importancia del espacio espectral para comprender los métodos de clasificación.
- ▶ El funcionamiento del algoritmo k-means de segmentación.
- ▶ El funcionamiento del algoritmo de máxima verosimilitud.



La vez pasada vimos

- ▶ Que a partir de esto podíamos definir la ρ_λ la firma espectral como una característica de cada cuerpo.
- ▶ La necesidad de definir categorías de uso y cobertura de forma concisa.
- ▶ La diferencia entre el concepto de categorías de uso y cobertura y clases espectrales.
- ▶ La importancia del espacio espectral para comprender los métodos de clasificación.
- ▶ El funcionamiento del algoritmo k-means de segmentación.
- ▶ El funcionamiento del algoritmo de máxima verosimilitud.



Esquema de presentación

Escenas del capítulo anterior

Filtrado

Técnicas pos-clasificación

Estimación de la precisión

Matriz de confusión

Estimación de áreas

Muestreo

Práctica



Filtrado

Nos va a permitir reducir algunos mitigar una limitación común en la clasificación como es la existencia de parches de escasa superficie. Suavizan las clasificaciones.



Filtrado

Nos va a permitir reducir algunos mitigar una limitación común en la clasificación como es la existencia de parches de escasa superficie. Suavizan las clasificaciones.



Técnicas pos-clasificación

4	4	4	4	-3	7
4	4	7	7	7	7
5	5	7	7	6	7
5	5	5	5	5	6
7	7	5	5	5	5
7	0	5	2		

 $=$

4	4	4	4	7	7
4	4	4	7	7	7
5	5	5	7	7	7
5	5	5	5	5	5
7	5	5	5	5	5
7	7	5	5		

Ejemplo de filtrado por mayoría.¹

¹filter.



Fusión

Nos permite convertir las clases de clasificación generadas por algún algoritmo en clases temáticas.



Técnicas pos-clasificación

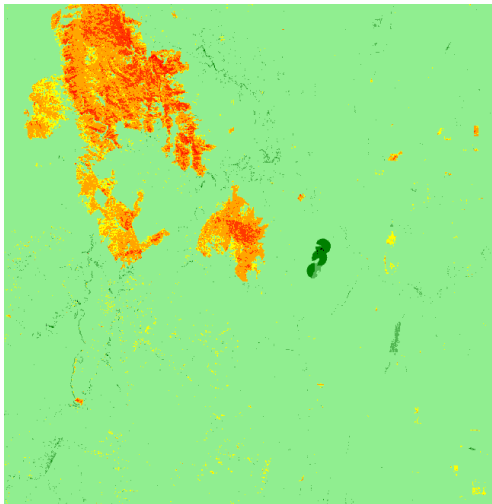


Imagen con clases fusionadas.



Técnicas pos-clasificación

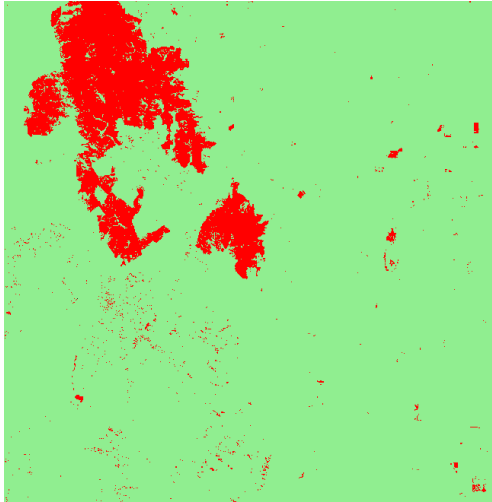


Imagen con clases fusionadas.



Técnicas pos-clasificación

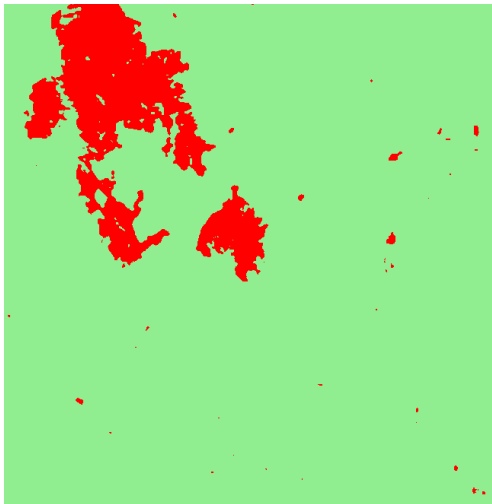


Imagen con clases fusionadas.



Esquema de presentación

Escenas del capítulo anterior

Filtrado

Técnicas pos-clasificación

Estimación de la precisión

Matriz de confusión

Estimación de áreas

Muestreo

Práctica

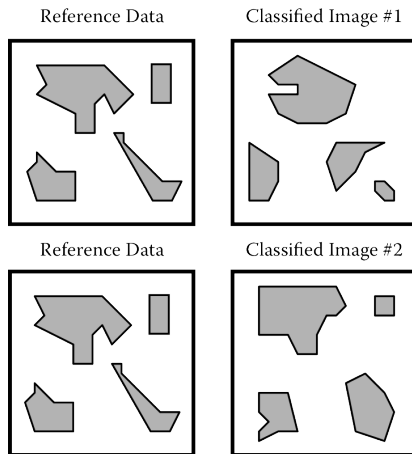


Objetivo de la validación

Lo que esperamos es asignarle a nuestra clasificación un cierto grado de confianza a partir de datos medidos en el terreno.



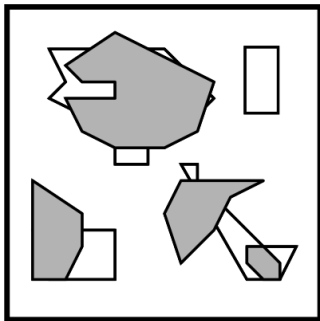
Técnicas pos-clasificación



Ejemplo de datos de referencia contra un mapa temático.²

²congalton2008assessing.



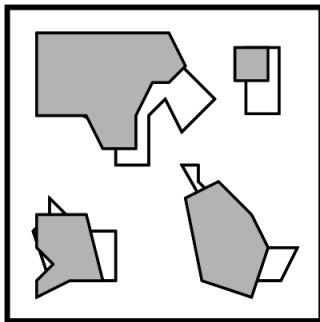


Comparación de área total.³

³congalton2008assessing.



Técnicas pos-clasificación



Comparación de espacial.⁴

⁴ congalton2008assessing.



Definición

Lo que esperamos es asignarle a nuestro mapa temático un cierto grado de precisión a partir de datos medidos en el terreno.



Matriz de confusión

Definición

$$\begin{bmatrix} & 1 & 2 & \dots & k & n_{i+} \\ 1 & n_{11} & n_{12} & \dots & n_{1k} & n_{1+} \\ 2 & n_{21} & n_{22} & \dots & n_{2k} & n_{2+} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots \\ k & n_{k1} & n_{k2} & \dots & n_{kk} & n_{k+} \\ n_{+j} & n_{+1} & n_{+2} & \dots & n_{+k} & N \end{bmatrix}$$



Definición

Donde

$$n_{i+} = \sum_j n_{ij}$$

$$n_{+j} = \sum_i n_{ij}$$

y donde n es el número total de muestras.



Ejemplo

Vamos a tomar sólo tres coberturas a modo de ejemplo

	<i>a</i>	<i>s</i>	<i>v</i>	
<i>a</i>	50	10	20	80
<i>s</i>	5	100	15	120
<i>v</i>	10	10	80	100
	65	120	115	300



Matriz de confusión

Presición total

$$O = \sum_i p_{ii}$$

Presición usuario

$$U_i = \frac{p_{ii}}{p_{i+}}$$

Presición productor

$$P_j = \frac{p_{jj}}{p_{+j}}$$



Matriz de confusión

Presición total

$$O = \sum_i p_{ii}$$

Presición usuario

$$U_i = \frac{p_{ii}}{p_{i+}}$$

Presición productor

$$P_j = \frac{p_{jj}}{p_{+j}}$$



Matriz de confusión

Presición total

$$O = \sum_i p_{ii}$$

Presición usuario

$$U_i = \frac{p_{ii}}{p_{i+}}$$

Presición productor

$$P_j = \frac{p_{jj}}{p_{+j}}$$



Matriz de confusión

Presición total

$$O = \sum_i p_{ii}$$

Presición usuario

$$U_i = \frac{p_{ii}}{p_{i+}}$$

Presición productor

$$P_j = \frac{p_{jj}}{p_{+j}}$$



Matriz de confusión

Fracción de la muestra

$$p_{ij} = \frac{W_i n_{ij}}{n_{i+}}$$

Probabilidad de j en los datos de campo

$$p_{+j} = \sum_i p_{ij}$$

Probabilidad de i en la clasificación

$$p_{i+} = \sum_j p_{ij}$$



Matriz de confusión

Fracción de la muestra

$$p_{ij} = \frac{W_i n_{ij}}{n_{i+}}$$

Probabilidad de j en los datos de campo

$$p_{+j} = \sum_i p_{ij}$$

Probabilidad de i en la clasificación

$$p_{i+} = \sum_j p_{ij}$$



Ejemplo

Si las áreas son $A_v = 1000$, $A_a = 500$ y $A_s = 500$ entonces

	a	s	v	
a	0,16	0,03	0,06	0,25
s	0,01	0,21	0,03	0,25
v	0,05	0,05	0,40	0,50
	0,23	0,29	0,49	0,77



Ejemplo

Si las áreas son $A_v = 1000$, $A_a = 500$ y $A_s = 500$ entonces

	a	s	v	
a	0,16	0,03	0,06	0,25
s	0,01	0,21	0,03	0,25
v	0,05	0,05	0,40	0,50
	0,23	0,29	0,49	0,77



Matriz de confusión

Matriz de confusión

Cualquier análisis sobre el error de una clasificación parte de la matriz de confusión.



Observación

Las áreas podemos estimarlas tanto a partir de p_{i+} y de p_{+j} .

- ▶ p_{i+} se conoce con certeza pero puede estar sesgado.
- ▶ p_{+j} presenta un sesgo menor pero debe ser estimado.



Observación

Las áreas podemos estimarlas tanto a partir de p_{i+} y de p_{+j} .

- ▶ p_{i+} se conoce con certeza pero puede estar sesgado.
- ▶ p_{+j} presenta un sesgo menor pero debe ser estimado.



Estimación de la precisión-Estimación de áreas

Podemos estimar la varianza de p_{+j} como

$$S(p_{+j}) = \sqrt{\frac{W_i p_{ij} - p_{ij}^2}{n_{i+} - 1}} \quad (1)$$

y por lo tanto

$$A_k = A_{total} \times p_{+j} \pm 1,96 \times A_{total} \times S(p_{+j})$$



Estimación de la precisión-Estimación de áreas

Podemos estimar la varianza de p_{+j} como

$$S(p_{+j}) = \sqrt{\frac{W_i p_{ij} - p_{ij}^2}{n_{i+} - 1}} \quad (1)$$

y por lo tanto

$$A_k = A_{total} \times p_{+j} \pm 1,96 \times A_{total} \times S(p_{+j})$$

.



Ejemplo

Las áreas con sus errores son

- ▶ $A_a = (433 \pm 88) km^2$
- ▶ $A_s = (579 \pm 70) km^2$
- ▶ $A_v = (988 \pm 95) km^2$



4 preguntas

1. ¿Qué categorías tengo?
2. ¿Qué unidad de muestreo usar?
3. ¿Cuántas muestras tomar?
4. ¿Cómo elegir las muestras?



¿Que categorías tengo?

Las clases tienen que ser

- ▶ Mutuamente exclusivas
- ▶ Totalmente exhaustivas

Además de tener un tamaño mínimo para ser considerado de esa clase.



¿Que categorías tengo?

Las clases tienen que ser

- ▶ Mutuamente exclusivas
- ▶ Totalmente exhaustivas

Además de tener un tamaño mínimo para ser considerado de esa clase.



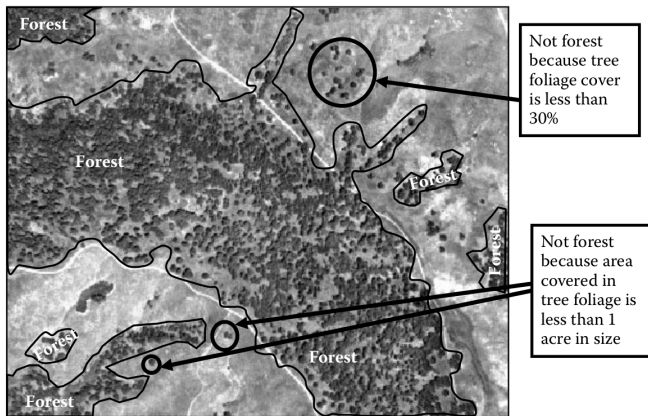
¿Que categorías tengo?

Las clases tienen que ser

- ▶ Mutuamente exclusivas
- ▶ Totalmente exhaustivas

Además de tener un tamaño mínimo para ser considerado de esa clase.





Clases de muestreo definidas en el terreno.⁵

⁵congalton2008assessing.



¿Qué unidad de muestreo usar?

- ▶ Un solo píxel.
- ▶ Un clúster de píxeles
- ▶ Un polígono
- ▶ Un clúster de polígonos



¿Qué unidad de muestreo usar?

- ▶ Un solo píxel.
- ▶ Un clúster de píxeles
- ▶ Un polígono
- ▶ Un clúster de polígonos



¿Qué unidad de muestreo usar?

- ▶ Un solo píxel.
- ▶ Un clúster de píxeles
- ▶ **Un polígono**
- ▶ Un clúster de polígonos



¿Qué unidad de muestreo usar?

- ▶ Un solo píxel.
- ▶ Un clúster de píxeles
- ▶ Un polígono
- ▶ Un clúster de polígonos



¿Como elegir las muestras?

- ▶ Al azar.
- ▶ Estratificado al azar.
- ▶ Sistemático.
- ▶ Clusters



¿Como elegir las muestras?

- ▶ Al azar.
- ▶ Estratificado al azar.
- ▶ Sistemático.
- ▶ Clusters



¿Como elegir las muestras?

- ▶ Al azar.
- ▶ Estratificado al azar.
- ▶ **Sistemático.**
- ▶ Clusters



¿Como elegir las muestras?

- ▶ Al azar.
- ▶ Estratificado al azar.
- ▶ Sistemático.
- ▶ Clusters



Logística

Todo lo que vimos va a estar supeditado a mi capacidad de realizar el muestreo.



Esquema de presentación

Escenas del capítulo anterior

Filtrado

Técnicas pos-clasificación

Estimación de la precisión

Matriz de confusión

Estimación de áreas

Muestreo

Práctica



Actividades prácticas de la sexta clase

1. Abrir las imágenes clasificadas y fusionadas por el método de clasificación supervisada y no supervisada.
2. Cargar los polígonos de validación correspondientes a cada clase.
3. Calcular la matriz de confusión correspondiente a cada clasificación.
4. Obtener la precisión global, del usuario, productor y el índice kappa.

