

Métricas de Evaluación

Ivan Cruz Aceves

ivan.cruz@cimat.mx

Centro de Investigación en Matemáticas, A.C. (CIMAT)

Febrero del 2019

Cubo- I304, Ext. 4506

Contenido

TIPOS DE CLASIFICACIÓN

MÉTRICAS DE EVALUACIÓN

Métricas alternativas

Etapas de Preprocesado en I.B.

Etapas de Postprocesado en I.B.

Tipos de clasificación

En el caso del problema de segmentación, un pixel puede ser categorizado en 4 eventos mutuamente diferentes:

Tipos de clasificación

En el caso del problema de segmentación, un pixel puede ser categorizado en 4 eventos mutuamente diferentes:

- ▶ **True Positive:** donde un pixel es clasificado correctamente como de interés. (1 en GT y 1 en S.I.)

Tipos de clasificación

En el caso del problema de segmentación, un pixel puede ser categorizado en 4 eventos mutuamente diferentes:

- ▶ **True Positive:** donde un pixel es clasificado correctamente como de interés. (1 en GT y 1 en S.I.)
- ▶ **True Negative:** donde un pixel es clasificado correctamente como de fondo. ($0 == 0$)

Tipos de clasificación

En el caso del problema de segmentación, un pixel puede ser categorizado en 4 eventos mutuamente diferentes:

- ▶ **True Positive:** donde un pixel es clasificado correctamente como de interés. (1 en GT y 1 en S.I.)
- ▶ **True Negative:** donde un pixel es clasificado correctamente como de fondo. (0 == 0)
- ▶ **False Negative:** donde un pixel es clasificado incorrectamente como de fondo. (1 == 0)

Tipos de clasificación

En el caso del problema de segmentación, un pixel puede ser categorizado en 4 eventos mutuamente diferentes:

- ▶ **True Positive:** donde un pixel es clasificado correctamente como de interés. (1 en GT y 1 en S.I.)
- ▶ **True Negative:** donde un pixel es clasificado correctamente como de fondo. (0 == 0)
- ▶ **False Negative:** donde un pixel es clasificado incorrectamente como de fondo. (1 == 0)
- ▶ **False Positive:** donde un pixel es clasificado incorrectamente como de interés. (0 == 1)

Tipos de clasificación

Las clasificaciones vistas anteriormente son ilustradas en la siguiente tabla:

—	Interés (GT)	Fondo (GT)
Interés (S.I.)	True Positive (TP)	False Positive (FP)
Fondo (S.I.)	False Negative (FN)	True Negative (TN)

Cuadro: Tabla de clasificación.

Obteniendo TP,FP,FN,TN

```
TP = sum(sum(img_ground == 1 & img_test == 1));  
FP = sum(sum(img_ground == 0 & img_test == 1));  
TN = sum(sum(img_ground == 0 & img_test == 0));  
FN = sum(sum(img_ground == 1 & img_test == 0));
```

Sensitivity

La métrica de Sensibilidad también conocida como **true positive rate**, se encarga de medir la proporción de positivos (interés) que fueron correctamente identificados por algún clasificador. Se encuentra definida en el intervalo $[0, 1]$, donde 0 es la peor clasificación y 1 es clasificación perfecta.

$$Sensitivity = \frac{TP}{(TP + FN)} \quad (1)$$

La expresión $(1 - Sensitivity)$, se conoce como **False negative rate**.

Specificity

La métrica de Especificidad también conocida como **true negative rate**, se encarga de medir la proporción de negativos (fondo) que fueron correctamente identificados por algún clasificador. Se encuentra definida en el intervalo $[0, 1]$, donde 0 es la peor clasificación y 1 es clasificación perfecta.

$$Specificity = \frac{TN}{(TN + FP)} \quad (2)$$

La expresión $(1 - Specificity)$, se conoce como **False positive rate**.

Accuracy

Esta métrica se encarga de medir la proporción de positivos (interés) y negativos (fondo) que fueron correctamente identificados por algún clasificador. Se encuentra definida en el intervalo $[0, 1]$, donde 0 es la peor clasificación y 1 es clasificación perfecta.

$$Accuracy = \frac{TP + TN}{(TP + FP + TN + FN)} \quad (3)$$

donde el denominador, en el caso de imágenes binarias puede ser sustituido por el **número total de píxeles** en la imagen procesada. Esta medida también puede conocerse como Misclassification error ($1 - Accuracy$).

Positive predictive value

Esta métrica se encarga de medir la proporción de positivos (interés) que fueron **correcta e incorrectamente** identificados por algún clasificador. Se encuentra definida en el intervalo $[0, 1]$, donde 0 es la peor clasificación y 1 es clasificación perfecta.

$$PPV = \frac{TP}{(TP + FP)} \quad (4)$$

La expresión $(1 - PPV)$, se conoce como **False discovery rate**.

Negative predictive value

Esta métrica se encarga de medir la proporción de negativos (fondo) que fueron **correcta e incorrectamente** identificados por algun clasificador. Se encuentra definida en el intervalo $[0, 1]$, donde 0 es la peor clasificación y 1 es clasificación perfecta.

$$NPV = \frac{TN}{(TN + FN)} \quad (5)$$

Matthews Correlation Coefficient

Esta medida de correlación es usada para valorar la calidad de una clasificación binaria tomando en consideración las razones de positivos y negativos como una medida de balance. Se encuentra definida en el intervalo $[-1, 1]$, donde -1 es la peor clasificación y 1 es clasificación perfecta.

$$MCC = \frac{((TP * TN) - (FP * FN))}{\sqrt{(TP + FP) * (TP + FN) * (TN + FP) * (TN + FN)}} \quad (6)$$

Jaccard Index

El coeficiente de Jaccard, también conocido como coeficiente de Tanimoto, mide la similitud contemplando positivos y negativos entre ambos conjuntos de datos. Representa una de las métricas más exactas en la literatura para evaluar resultados contra un ground-truth real. Se encuentra definido en el intervalo $[0, 1]$, donde 0 es la peor clasificación y 1 es clasificación perfecta.

$$Jaccard = \frac{TP}{FP + FN + TP} \quad (7)$$

Dice coefficient

El coeficiente de Dice esta **directamente relacionado con el coeficiente de Jaccard**, por lo cual ambos pueden ser utilizados indistintamente en problemas de clasificación. En general, se prefiere usar este coeficiente en lugar de Jaccard. De igual forma, se encuentra definido en el intervalo $[0, 1]$, donde 0 es la peor clasificación y 1 es clasificación perfecta.

$$Dice = \frac{2TP}{FP + FN + 2TP} \quad (8)$$

Métricas alternativas

- ▶ Root-mean-square error
- ▶ Hausdorff distance
- ▶ Absolute mean brightness error
- ▶ Peak signal-to-noise ratio
- ▶ Baddeleys delta metric
- ▶ Edge mismatch
- ▶ Region nonuniformity (no ground-truth)

Métricas alternativas

- ▶ Root-mean-square error
- ▶ Hausdorff distance
- ▶ Absolute mean brightness error
- ▶ Peak signal-to-noise ratio
- ▶ Baddeleys delta metric
- ▶ Edge mismatch
- ▶ Region nonuniformity (no ground-truth)

Una vez calculadas varias métricas, puede usarse la media como medida de desempeño global o un método de ranqueo.

Por ejemplo ranqueo promediado: para cada imagen se rankean los algoritmos desde $\{1, \dots, N\}$ del promedio de los criterios, para al final promediar rankeos.

Mostrando resultados¹

Table 3

Comparative analysis of the proposed GMF-DE method with six state-of-the-art vessel segmentation methods using the test set of 40 angiograms.

Segmentation method	Sensitivity	Specificity	Accuracy	Average
Kang et al. [18]	0.5113	0.9863	0.9565	0.8180
Qian et al. [6]	0.8355	0.8521	0.8511	0.8462
Wang et al. [12]	0.6217	0.9753	0.9532	0.8501
Chanwimaluang and Fan [17]	0.8647	0.8721	0.8717	0.8695
Al-Rawi et al. [22]	0.8492	0.8818	0.8798	0.8703
Kang et al. [20]	0.8359	0.9110	0.9063	0.8844
Proposed GMF-DE	0.8247	0.9620	0.9534	0.9134

¹Applied Soft Computing. On the performance of nature inspired algorithms for the automatic segmentation of coronary arteries using Gaussian matched filters.46, 665-676.(2016)

Mostrando resultados²

Table 1 Comparative analysis of the evolutionary thresholding method with five state-of-the-art thresholding methods using the training set of 20 angiograms

Thresholding method	Sensitivity	Specificity	Accuracy	PPV	NPV	Average
Proposed method	0.7703	0.9764	0.9664	0.6256	0.9881	0.8654
Pal and Pal [19]	0.8171	0.9690	0.9616	0.5739	0.9904	0.8624
RATS [20, 21]	0.3646	0.9965	0.9657	0.8410	0.9684	0.8272
Otsu [18]	0.3144	0.9976	0.9643	0.8682	0.9660	0.8221
Moments [22]	0.2434	0.9986	0.9618	0.8958	0.9627	0.8125
Kapur et al. [23]	0.2126	0.9988	0.9605	0.8995	0.9612	0.8065

²Segmentation of coronary angiograms using a vesselness measure and evolutionary thresholding, in Book: Design of intelligent systems based on fuzzy logic, neural networks and nature-inspired optimization from Studies in Computational Intelligence,(601), Springer, 2015

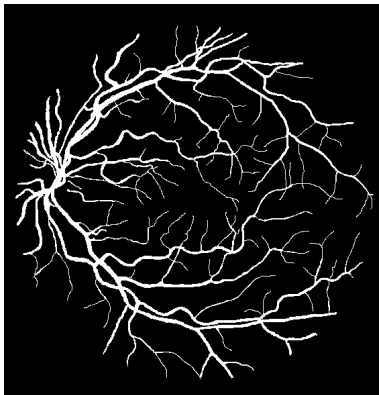
Etapa de preprocesado

Delimitación de área de interés.



Etapa de preprocesado

Delimitación de área de interés.



Etapa de postprocesado

Length filtering



Figura: (a) Img. Original, (b) resultado (bwareaopen, 1000).

Rellenado de orificios

Cerradura (Closing)

Operación compuesta de dilatación y erosión.

$$A = I \bullet S = (I \oplus S) \ominus S$$

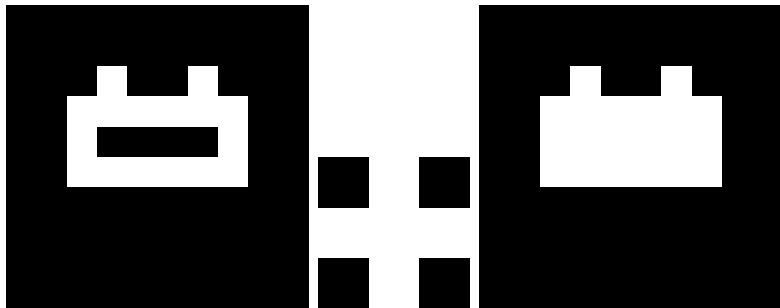


Figura: (a) Img. Original, (b) SE, (c) resultado.

Tarea 2 para entregar

- ▶ Del Survey platicado en clase:
- ▶ Implementar un método de umbralización del EdA.
- ▶ Analizar su desempeño sobre una BD.