



Tecnológico de Monterrey

Instituto Tecnológico y de Estudios Superiores de Monterrey

Análisis de Sistemas de Imagenología (Grupo 201 y 202)

Reto - Semana 3 (Bloque 2)

Profesor

Dr. José Gerardo Tamez Peña

Equipo Pegaso:

Daniela Alejandra Valdes Torres	A00832312
Diego De La Barrera Martínez	A01197739
Alexa María de León Durán	A01382990
Juan Luis Flores Sánchez	A01383088
Azul Sofía Moctezuma Enriquez	A01562585

21 de abril de 2023

Reto - Semana 3 (Bloque 2)

1. Introducción

Las Medidas de Comparación de Segmentación son una herramienta esencial para evaluar la calidad de las segmentaciones de imágenes. Para la elección de la medida de comparación adecuada se debe de tomar en cuenta el objetivo de la segmentación y de las características de las imágenes que se están segmentando. Sin embargo, es importante considerar que cada una de estas medidas tiene sus propias ventajas y desventajas, por lo que la elección de cada medida dependerá del tipo de imágenes que se busque segmentar y del objetivo del análisis.

2. Objetivo

El objetivo es conocer las medidas de comparación de segmentación existentes con la finalidad de encontrar la referencia de segmentación que más se adapte a las segmentaciones que realizaremos en futuras entregas.

3. Marco Teórico

3.1 Medidas de Comparación de Segmentación

En la siguiente tabla se presentan definiciones importantes para comprender los siguientes conceptos :

Siglas	Definición
VP	Datos identificados en la clase “Positivos” y que con base en el conjunto de datos brindado se corrobora que son “Positivos”
VN	Datos identificados en la clase “Negativos” y que con base en el conjunto de datos brindado se corrobora que son “Negativos”
FP	Datos identificados en la clase “Positivos” y que con base en el conjunto de datos brindado se corrobora que son “Negativos”
FN	Datos identificados en la clase “Negativos” y que con base en el conjunto de datos brindado se corrobora que son “Positivos”
(Microsoft, 2023)	

- **Exactitud:** La exactitud en análisis de imágenes se refiere a qué tan cercanos están los resultados del análisis de imágenes a la verdad de referencia. En otras palabras, la exactitud mide la capacidad del algoritmo de análisis de imágenes para identificar correctamente los objetos de interés en una imagen y excluir otros elementos no relevantes (Ye, Gilmore & Rakshit, 2018). Para realizar dicha evaluación de exactitud, es necesario utilizar un dataset de referencia para determinar la exactitud del resultado clasificado. La exactitud se representa de

0 a 1, de modo que 1 equivale a una exactitud del 100% (ARCGis, 2022), siendo obtenida a partir de la ecuación (1).

$$Exactitud = \frac{VP+VN}{VP+FP+FN+VN} \quad (1)$$

- **Precisión:** La precisión en análisis de imágenes es una medida de cuántos de los objetos identificados por el algoritmo son realmente objetos de interés. En otras palabras, mide la capacidad del algoritmo para detectar correctamente los objetos de interés (VP, Verdaderos Positivos) sin incluir objetos que no son relevantes (VN, Verdaderos Negativos) (Dang, 2022). Esta medición se obtiene con base en la ecuación (2).

$$Precisión = \frac{VP}{VP+FP} \quad (2)$$

- **Sensibilidad:** La sensibilidad en análisis de imágenes es una medida estadística para evaluar la calidad de la imagen, esta se calcula como la fracción de verdaderos positivos (TP) en relación con el total de píxeles positivos (TP + FN) donde TP son los pixeles que pertenecen a la región segmentada y que también están presentes en la máscara de verdad terreno y FN son los pixeles que pertenecen a la región segmentada pero no se encuentran presentes en la máscara de verdad de terreno. Matemáticamente la sensibilidad se define como (3):

$$Sensibilidad = \frac{VP}{(VP+FN)} \quad (3)$$

Dada la ecuación de sensibilidad, cuando los valores de sensibilidad son altos indica que los píxeles de la máscara de verdad terreno incluyen una buena segmentación por el algoritmo indicando una buena calidad de la segmentación. Esta es una medida importante que se utiliza junto con otras medidas como la especificidad y el índice de Jaccard que evalúan la precisión y calidad de la segmentación generada por el algoritmo.

- **Especificidad:** Esta medida de comparación es importante en la evaluación de la calidad de las segmentaciones de imágenes ya que indica la proporción de píxeles o voxels que son clasificados correctamente como fondo por el algoritmo de segmentación, es decir, cuántos píxeles que en realidad no forman parte de la estructura de interés son identificados correctamente como tal.

Para realizar su cálculo se pueden realizar las mismas métricas para la evaluación de la segmentación de imágenes, pero realizando un enfoque en la clasificación del fondo, es decir, calculando la tasa de verdaderos negativos (VN).

$$VN = \frac{n \text{ de píxeles clasificados correctamente como fondo.}}{n \text{ de píxeles que pertenecen al fondo total en la máscara de referencia.}}$$

Una vez calculada la tasa se realiza la división del VN entre el número total de píxeles que se clasificaron como fondo (4), siendo FP el número de falsos positivos (Sarmiento, s.f):

$$\text{Especificidad} = \frac{VN}{(VN + FP)} \quad (4)$$

Un valor de especificidad alto nos indica que el algoritmo es capaz de realizar una buena distinción entre la estructura de interés de fondo y que clasifica correctamente la mayoría de píxeles que pertenecen al fondo. (Molina, 2011)

- **Jaccard:** El índice de similitud Jaccard es utilizado para comparar la similitud y diversidad de dos conjuntos. Para calcularla se utiliza la proporción entre el tamaño de la intersección de los conjuntos. Este índice nos permite evaluar la calidad de las segmentaciones en las imágenes, medir la segmentación generada por un algoritmo y una referencia o máscara. Los píxeles en las imágenes se dividen en dos clases: los que pertenecen a la sección segmentada y los que no pertenecen a ella.

Este índice se calcula como la fracción de píxeles que coinciden en ambas segmentaciones en relación con el total de píxeles de ambas segmentaciones midiendo la similitud y su valor oscila de 0 y 1 donde el valor de 1 nos indica que las segmentaciones son idénticas.

La fórmula del índice de Jaccard es la siguiente:

$$J(A, B) = |A \cap B| / |A \cup B|$$

Donde A y B representan dos conjuntos de elementos, esto es utilizado en el campo del procesamiento de imágenes y la visión por computadora para evaluar la calidad de estas segmentaciones.

- **Dice similarity y Overlap measure:** Un índice de superposición espacial simple es el coeficiente de similitud de Dice (DSC), propuesto por primera vez por Dice. El coeficiente de similitud de Dice es un índice de superposición espacial y una métrica de validación de reproducibilidad. Fleiss también la llamó proporción de concordancia específica. El valor de un DSC varía de 0, lo que indica que no hay superposición espacial entre dos conjuntos de resultados de segmentación binaria, a 1, lo que indica una superposición completa. Se ha adoptado el coeficiente de similitud de Dice para validar la segmentación de lesiones de sustancia blanca en resonancias magnéticas y la zona periférica de la glándula prostática en braquiterapia prostática.

La principal métrica de validación del índice de superposición espacial es DSC. El DSC mide la superposición espacial entre dos segmentaciones, las regiones objetivo A y B, y se define como $DSC(A,B)=2(A \cap B)/(A+B)$ donde \cap es la intersección. En la segmentación manual binaria, este coeficiente puede derivarse de una tabla de contingencia de probabilidades de clasificación de segmentación de dos por dos. Conceptualmente, DSC también es un caso especial de la estadística kappa comúnmente utilizada en el análisis de confiabilidad, cuando hay una cantidad mucho mayor de vóxeles de fondo que la de los vóxeles objetivo (Zou, et al. 2004).

- **Kappa:** La estadística kappa se usa con frecuencia para probar la confiabilidad entre evaluadores. La importancia de la confiabilidad del evaluador radica en el hecho de que representa hasta qué punto los datos recopilados en el estudio son representaciones correctas de las variables medidas. La medición de la extensión en que los recopiladores de datos (evaluadores) asignan la misma puntuación a la misma variable se denomina confiabilidad entre evaluadores. Si bien ha habido una variedad de métodos para medir la confiabilidad entre evaluadores, tradicionalmente se medía como “Percent agreement”, calculado como el número de “Agreement scores” dividido por el número total de puntajes.

Muchas situaciones en la industria de la salud dependen de varias personas para recopilar datos de investigación o de laboratorio clínico. La cuestión de la coherencia o el acuerdo entre las personas que recopilan datos surge debido a la variabilidad entre los observadores humanos. Por lo tanto, los estudios de investigación bien diseñados deben incluir procedimientos que midan el acuerdo entre los diversos recolectores de datos. Los diseños de estudio generalmente implican capacitar a los recolectores de datos y medir hasta qué punto registran los mismos puntajes para los mismos fenómenos. Rara vez se logra un acuerdo perfecto, y la confianza en los resultados del estudio es en parte una función de la cantidad de desacuerdo o error introducido en el estudio debido a la inconsistencia entre los recolectores de datos. El grado de acuerdo entre los recolectores de datos se denomina "confiabilidad entre evaluadores" (McHugh, 2012).

3.2 Códigos de MATLAB para Comparar Segmentaciones

- **Código 1: Función que Acepte la Segmentación Realizada como Entrada y Cree un Archivo con 3 Clases (Materia Gris, Materia Blanca, Fondo/Otros):**

```
%% BI2007B Análisis de Sistemas de Imagenología
%% Equipo PEGASO
%% Clasificación e Importación de Segmentación en 3 Clases (Materia Gris,
Materia Blanca, Fondo/Otros)

function create_three_class_file(input_file, output_file)
    % Leer los datos del archivo de etiquetas usando la función read_label
de FreeSurfer
    [~, labels, colormap] = read_label(input_file);

    % Crear una matriz con etiquetas de tres clases: Materia Gris, Materia
Blanca y Otros
    three_class_labels = zeros(size(labels));
    gray_matter_indices = colormap.table(:,5) == 3 | colormap.table(:,5) ==
42 | colormap.table(:,5) == 8 | colormap.table(:,5) == 47;
```

```

    white_matter_indices = colormap.table(:,5) == 2 | colormap.table(:,5) ==
7 | colormap.table(:,5) == 16 | colormap.table(:,5) == 28 |
colormap.table(:,5) == 46;
    three_class_labels(gray_matter_indices) = 1;
    three_class_labels(white_matter_indices) = 2;
    three_class_labels(~(gray_matter_indices | white_matter_indices)) = 3;

    % Escribir los datos de etiquetas de tres clases en el archivo de salida
    write_label(three_class_labels, output_file);
end

```

- Código 2: Código que Compare 2 Segmentaciones para Calcular Exactitud, Precisión, Sensibilidad, Especificidad, Coeficiente de Jaccard, Dice Similarity, Medida de Traslapo (Overlap Measure) y Kappa:

```

%% BI2007B Análisis de Sistemas de Imagenología
%% Equipo PEGASO
%% Comparación de Segmentaciones para la Obtención de Métricas Comparativas

% Cargar las dos segmentaciones realizadas con FreeSurfer
seg1 = MRIread('segmentacion1.mgz');
seg2 = MRIread('segmentacion2.mgz');
% Obtener la matriz de confusión
confusionMatrix = zeros(2, 2);
confusionMatrix(1, 1) = sum(seg1.vol == 1 & seg2.vol == 1);
confusionMatrix(1, 2) = sum(seg1.vol == 1 & seg2.vol == 0);
confusionMatrix(2, 1) = sum(seg1.vol == 0 & seg2.vol == 1);
confusionMatrix(2, 2) = sum(seg1.vol == 0 & seg2.vol == 0);
% Calcular las métricas de evaluación
accuracy = (confusionMatrix(1, 1) + confusionMatrix(2, 2)) /
sum(confusionMatrix(:));
precision = confusionMatrix(1, 1) / (confusionMatrix(1, 1) +
confusionMatrix(1, 2));
sensitivity = confusionMatrix(1, 1) / (confusionMatrix(1, 1) +
confusionMatrix(2, 1));
specificity = confusionMatrix(2, 2) / (confusionMatrix(2, 2) +
confusionMatrix(1, 2));
jaccard = confusionMatrix(1, 1) / (confusionMatrix(1, 1) +
confusionMatrix(1, 2) + confusionMatrix(2, 1));

```

```

dice = 2 * confusionMatrix(1, 1) / (2 * confusionMatrix(1, 1) +
confusionMatrix(1, 2) + confusionMatrix(2, 1));
overlap = confusionMatrix(1, 1) / (confusionMatrix(1, 1) +
min(confusionMatrix(1, 2), confusionMatrix(2, 1)));
kappa = (accuracy - sum(sum(confusionMatrix))*sum(sum(confusionMatrix(:,
:)))/(sum(sum(confusionMatrix(1, :)))*sum(sum(confusionMatrix(:, 1)))/(1
-
sum(sum(confusionMatrix))*sum(sum(confusionMatrix(:,
:)))/(sum(sum(confusionMatrix(1, :)))*sum(sum(confusionMatrix(:, 1))));
% Mostrar los resultados en pantalla
disp(['Exactitud: ', num2str(accuracy)]);
disp(['Precisión: ', num2str(precision)]);
disp(['Sensibilidad: ', num2str(sensitivity)]);
disp(['Especificidad: ', num2str(specificity)]);
disp(['Coeficiente de Jaccard: ', num2str(jaccard)]);
disp(['Similarity Dice: ', num2str(dice)]);
disp(['Overlap measure: ', num2str(overlap)]);
disp(['Kappa: ', num2str(kappa)]);

```

4. Referencias

- [1]. ARCGis. (2022). Descripción general de la clasificación de imágenes. Abril 14, 2023. Recuperado de: <https://pro.arcgis.com/es/pro-app/latest/help/analysis/image-analyst/overview-of-image-classification.htm>
- [2]. Dang, T. (2022). Guide to accuracy, precision, and recall. Abril 21, 2023. Recuperado de: <https://www.mage.ai/blog/definitive-guide-to-accuracy-precision-recall-for-product-developers>
- [3]. McHugh, M. L. (2012). *Interrater Reliability: The kappa statistic*. Biochemia medica. Retrieved April 21, 2023, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3900052/>
- [4]. Microsoft. (2023). Characteristics and limitations for using Image Analysis. Abril 21, 2023. Recuperado de: <https://learn.microsoft.com/en-us/legal/cognitive-services/computer-vision/imageanalysis-characteristics-and-limitations>
- [5]. Molina-Cortés, J., Restrepo-Martínez, A., & Branch-Bedoya, J. W. (2011, December). *Optimización de la segmentación local de Sauvola Aplicada a la detección de defectos superficiales en escenas con iluminación no homogénea*. SciELO. Retrieved April 21, 2023, from http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0123-77992011000200004
- [6]. Sarmiento, A., Fondón, I., Velasco, M., Qaisar, A., Aguilera, P. (s.f.). Modelo de Mezcla de Gaussianas Generalizadas para Segmentación de Melanomas. Retrieved April 21, 2023, from https://www.researchgate.net/profile/Pablo-Aguilera-6/publication/266910590_Modelo_de_Mezcla_de_Gaussianas_Generalizadas_para_Segmentacion_de_Melanomas/links/547f05650cf2c1e3d2dc3933/Modelo-de-Mezcla-de-Gaussianas-Generalizadas-para-Segmentacion-de-Melanomas.pdf
- [7]. Ye, S. Gilmore, R. & Rakshit, R. (2018). A review of accuracy assessment for object-based image analysis: From per-pixel to per-polygon approaches. Abril 21, 2023. Recuperado de: <https://doi.org/10.1016/j.isprsjprs.2018.04.002>
- [8]. Zou, K. H., Warfield, S. K., Bharatha, A., Tempany, C. M. C., Kaus, M. R., Haker, S. J., Wells, W. M., Jolesz, F. A., & Kikinis, R. (2004, February). *Statistical validation of image segmentation quality based on a spatial overlap index*. Academic radiology. Retrieved April 21, 2023, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1415224/#:~:text=Dice%20similarity%20coefficient%20is%20a,to%201%2C%20indicating%20complete%20overlap>.