

CENTRO DE INVESTIGACIÓN EN
COMPUTACIÓN

CLASIFICACIÓN INTELIGENTE DE PATRONES

Matriz de confusión: medidas de desempeño

Autor:
Diego Noguez Ruiz

Profesor:
Dr. Cornelio Yáñez
Márquez

1 de Abril 2023



Índice general

Índice general	1
1. Introducción	3
2. Desarrollo	3
2.1. Propósito de la tarea	3
2.2. Parte 1:	3
2.3. 1.a)	3
2.4. 1.b)	3
2.5. 1.c)	4
2.6. 1.d)	4
2.7. Parte 2:	5
2.8. 2.a)	5
2.9. 2.b)	5
2.10. 1.c)	5
2.11. 1.d)	5
2.12. Parte 3:	6
2.13. 3.a)	6
2.14. 3.b)	6
2.15. 3.c)	6
2.16. 3.d)	6
2.17. Parte 4:	6
2.18. 4.a)	7
2.19. 4.b)	7
2.20. 4.c)	7
2.21. 4.d)	7
3. Conclusiones y trabajo futuro	7

Índice de figuras

Figura 1 - Datasets usados en el artículo	4
---	---

1. Introducción

Durante las últimas sesiones, se definió un concepto para estudiar las medidas de desempeño, la matriz de confusión. Se analizaron sus elementos y se utilizó para el cálculo del *Accuracy*. Sin embargo, es importante ampliar nuestro conocimiento en lo referente a las medidas de desempeño. Por lo tanto, en el RD 07 se presentaron tres nuevas medidas: *Sensitivity*, *Specificity* y *Balanced Accuracy*.

En esta tarea, se buscarán artículos indexados en JCR Q1 que hagan uso de las medidas de desempeño estudiadas con la finalidad de concluir que tan usual es encontrarlas en la literatura.

2. Desarrollo

2.1. Propósito de la tarea

Investigar la presencia de las medidas de desempeño en artículos recientes publicados en revistas JCR.

2.2. Parte 1:

Buscar un artículo reciente en revista JCR Q1, donde se use la medida de desempeño accuracy.

2.3. 1.a)

Instrucciones: Reportar la ficha bibliográfica del artículo.

J. Sun, G. Zhong, K. Huang, and J. Dong, “Banzhaf random forests: Cooperative game theory based random forests with consistency,” *Neural Networks*, vol. 106, pp. 20–29, 2018. [1]

2.4. 1.b)

Instrucciones: Reportar el nombre del dataset y los nombres de las clases.

En este artículo aplican distintos clasificadores a 12 datasets de UCI ML, adicionalmente se usaron 3 datasets del mundo real, ver fig. 1

Data sets	No.examples	No.features	No.classes
ionosphere	35	34	2
wine	178	13	3
sonar	208	20	2
housing	506	13	2
dermatology	366	34	6
pima	768	8	2
vehicle	846	18	4
waveform40	5000	40	3
newsgroups	16242	100	4
satimage	6435	36	6
musk2	6598	166	2
shuttle	14516	9	7
usps	9298	256	10
Yale	165	1024	15
isolet	6238	617	26

Figura 1: Datasets usados en el artículo

Uno de los datasets es el *ionosphere* [2], el cual es biclase las cuales son:

- g: Good
- b: Bad

2.5. 1.c)

Instrucciones: Reportar el valor de IR del dataset.

- La clase b tiene 126 patrones
- La clase g tiene 225 patrones

Entonces se concluye que

$$IR = \frac{225}{126} = 1.785714$$

En este caso tenemos un dataset desbalanceado.

2.6. 1.d)

Instrucciones: Verificar si es adecuado, o no, utilizar accuracy. Explicar brevemente.

No es adecuado en base a lo estudiado, debido a que se definió que era adecuado sii el dataset es balanceado (IR menor o igual a 1.5).

2.7. Parte 2:

Buscar un artículo reciente en revista JCR Q1, donde se use la medida de desempeño sensitivity (recall)

2.8. 2.a)

Instrucciones: Reportar la ficha bibliográfica del artículo.

Q.-F. Zhou, H. Zhou, Y.-P. Ning, F. Yang, and T. Li, “Two approaches for novelty detection using random forest,” Expert Systems with Applications, vol. 42, no. 10, pp. 4840–4850, 2015. 7 [3]

2.9. 2.b)

Instrucciones: Reportar el nombre del dataset y los nombres de las clases.

En este artículo se aplican varios clasificadores a un dataset artificial y a cinco datasets de UCI. En particular, se utiliza el dataset Wine [4], que consta de tres clases que representan diferentes viñedos de donde proviene el vino. Estas clases se identifican con los números 1, 2 y 3.

2.10. 1.c)

Instrucciones: Reportar el valor de IR del dataset.

- Clase 1: 59
- Clase 2: 71
- Clase 3: 48

Entonces se concluye que

$$IR = \frac{71}{48} = 1.479167$$

En este caso tenemos un dataset balanceado debido a que $IR < 1.5$.

2.11. 1.d)

Instrucciones: Justificar brevemente el uso de sensitivity.

El *sensitivity* es equivalente al *Accuracy de la clase positiva*, y es correcto hacer uso de esta medida sin importar si el dataset es balanceado o no. Usando la notación de la matriz de confusión esta medida de desempeño se define de la siguiente manera.

Definición 1.

$$Recall = \frac{TP}{TP + FN}$$

2.12. Parte 3:

Buscar un artículo reciente en revista JCR Q1, donde se use la medida de desempeño specificity

2.13. 3.a)

Instrucciones: Reportar la ficha bibliográfica del artículo.

M. A. Chikh, M. Saidi, and N. Settouti, “Diagnosis of Diabetes Diseases Using an Artificial Immune Recognition System2 (AIRS2) with Fuzzy K- nearest Neighbor,” Journal of Medical Systems, vol. 36, pp. 2721–2729, Oct. 2012. [5]

2.14. 3.b)

Instrucciones: Reportar el nombre del dataset y los nombres de las clases.

El dataset con el que trabajan es el *Pima Indians Diabetes Database* [6]. Las clases son 0 y 1 las cuales representan sano y diabético respectivamente.

2.15. 3.c)

Instrucciones: Reportar el valor de IR del dataset.

- Clase 0 (Sanos): 262
- Clase 1 (Diabetes): 130

Entonces se concluye que

$$IR = \frac{262}{130} = 2.015385$$

Por lo tanto el dataset es desbalanceado.

2.16. 3.d)

Instrucciones: Justificar brevemente el uso de specificity.

El *specificity* es equivalente al *Accuracy de la clase negativa*, y es correcto hacer uso de esta medida sin importar si el dataset es balanceado o no. Usando la notación de la matriz de confusión esta medida de desempeño se define de la siguiente manera.

Definición 2.

$$Specificity = \frac{TN}{TN + FP}$$

2.17. Parte 4:

Buscar un artículo reciente en revista JCR Q1, donde se use la medida de desempeño balanced accuracy

2.18. 4.a)

Instrucciones: Reportar la ficha bibliográfica del artículo.

P. Bergmeir, C. Nitsche, J. Nonnast, and M. Bargende, “Classifying component failures of a hybrid electric vehicle fleet based on load spectrum data,” *Neural Computing and Applications*, vol. 27, pp. 2289–2304, Nov. 2016. 8 [7]

2.19. 4.b)

Instrucciones: Reportar el nombre del dataset y los nombres de las clases.

En el artículo, los autores mencionan el *load spectrum data*, pero no especifican de donde se pueden obtener los datos. Es posible que estos sean privados. Con respecto a las clases, mencionan que hay dos: autos defectuosos y no defectuosos.

2.20. 4.c)

Instrucciones: Reportar el valor de IR del dataset.

No es posible calcular el IR ya que no se tiene acceso al dataset. Sin embargo, los autores mencionan que existe un desbalance de clases, ya que hay más patrones pertenecientes a vehículos no defectuosos en comparación con los defectuosos.

2.21. 4.d)

Instrucciones: Justificar brevemente el uso de balanced accuracy.

El *balanced accuracy* es una medida de desempeño útil, en particular para ocasiones en las que se está trabajando con datasets desbalanceados como es el caso. Se define de la siguiente manera.

Definición 3.

$$\text{Balanced Accuracy} = \frac{\text{Recall} + \text{Specificity}}{2}$$

3. Conclusiones y trabajo futuro

Indudablemente las medidas de desempeño estudiadas son ampliamente usadas en artículos indexados JCR Q1, lo cual, es prueba de que son medidas importantes; sin embargo, la búsqueda del artículo referente a *Balanced Accuracy* me ha sido complicada, debido a que no encontraba un artículo que cumpliera dichas especificaciones. Por otro lado, me ha causado sorpresa que en el artículo de la parte 1, se haya calculado el *Accuracy*, ya que se estaba trabajando con un dataset desbalanceado, lo cual nos puede hacer dudar de la calidad del artículo y de la revista.

En breve, se nos ha notificado que estudiaremos otra medida de desempeño, para así terminar de conocer las medidas más conocidas en el estado del arte.

Referencias

- [1] J. Sun, G. Zhong, K. Huang, and J. Dong, “Banzhaf random forests: Cooperative game theory based random forests with consistency,” *Neural Networks*, vol. 106, pp. 20–29, 2018.
- [2] UCI ML, “Ionosphere data set.” <https://archive.ics.uci.edu/ml/datasets/ionosphere>. Visitado en Marzo 2023.
- [3] Q.-F. Zhou, H. Zhou, Y.-P. Ning, F. Yang, and T. Li, “Two approaches for novelty detection using random forest,” *Expert Systems with Applications*, vol. 42, no. 10, pp. 4840–4850, 2015.
- [4] UCI ML, “Wine data set.” <https://archive.ics.uci.edu/ml/datasets/wine>. Visitado en Marzo 2023.
- [5] M. A. Chikh, M. Saidi, and N. Settouti, “Diagnosis of Diabetes Diseases Using an Artificial Immune Recognition System2 (AIRS2) with Fuzzy K-nearest Neighbor,” *Journal of Medical Systems*, vol. 36, pp. 2721–2729, Oct. 2012.
- [6] National Institute of Diabetes and Digestive and Kidney Diseases//Kaggle, “Pima indians diabetes database.” <https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database>. Visitado en Marzo 2023.
- [7] P. Bergmeir, C. Nitsche, J. Nonnast, and M. Bargende, “Classifying component failures of a hybrid electric vehicle fleet based on load spectrum data,” *Neural Computing and Applications*, vol. 27, pp. 2289–2304, Nov. 2016.