Lección 5. Métricas de evaluación de los métodos de detección de anomalías

UCO ONLINE

La detección de anomalía

Métricas de evaluación de los métodos de detección de anomalías

La detección de anomalías necesita ser evaluada para poder comprobar si los algoritmos o modelos están funcionando bien. Por ello, partimos de que en el conjunto que se va a evaluar se conocen las anomalías. Esto se hará con independencia de que el algoritmo trabaje como un modelo no supervisado y, por tanto, no tenga dicha información para generar el modelo. En la fase de evaluación, consideraremos que tenemos esta información, para poder aplicar las métricas que aquí se estudian y evaluar los modelo. Si no se dispone de dicha información, la evaluación debería basarse en métricas que valoren el agrupamiento de los datos, tal y como se ha visto también en esta asignatura.

Las características de estos problemas, en los que se tienen muchos menos ejemplos de anomalías comparados con los ejemplos que son normales, hace que los modelos no deban medirse con una exactitud global, ya que nos daría por buenos algoritmos que realmente no detectan ninguna anomalía. Esto es debido al desbalanceo entre las clases. Debido a ello es necesario usar métricas que tengan en cuenta el desbalanceo.

En esta sección se definen las métricas que frecuentemente son usadas para evaluar el rendimiento de los modelos de los métodos de detección de anomalía [1,2].

4.1 Matriz de confusión

La matriz de confusión es una herramienta que nos muestra el desempeño del algoritmo, describiendo cómo se distribuyen los valores reales y nuestras predicciones mediante 4 valores basados en los valores reales y predichos [3].

En el problema de anomalía con el que vamos a trabajar, se van a considerar dos clases, la clase anónala (que es tratada como la clase positiva, +) y la clase normal (que es tratada como la clase negativa, -). Después de aplicar el modelo a nuestros datos obtendremos los cuatro valores que se muestran a continuación y que pertenecen a la matriz de confusión (tabla 1):

- Verdaderos positivos (TP): número de elementos cuya clase real es la anómala y el modelo los ha determinado como anómalos.
- Falsos positivos (FP): número de elementos cuya clase real es normal y el modelo los ha determinado como anómalos.
- Verdaderos negativos (TN): número de elementos cuya clase real es normal y el modelo las ha determinado como normales.
- Falsos negativos (FN): número de elementos cuya clase real es anómala y el modelo las ha determinado como normales.

		Predicción	
		Anomalía (+)	Normal (-)
Real	Anomalía (+)	Verdaderos positivos (TP)	Falsos negativos (FN)
	Normal (-)	Falsos positivos (FP)	Verdaderos negativos (TN)

Tabla 1. Elementos de la matriz de confusión

4.2 Sensibilidad

En inglés se conoce como *Recall* o *Sensibility*, también se conoce como Tasa de Verdaderos Positivos (TPR, *True Positive Rate*). Es la proporción de casos positivos que fueron correctamente identificadas por el algoritmo. Por tanto, miden la cantidad de anomalías que detecta el método. Nos da una visión global tanto de cómo de bien detecta las anomalías, como del número de falsos negativos que se obtienen [4].

Se calcula mediante la siguiente ecuación, donde utilizamos los valores de la matriz de confusión:

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

4.3 Especificidad

En inglés conocido como Especificity y también conocido como Tasa de Verdaderos Negativos, (TNR, True Negative Rate). Se trata de los casos negativos que el algoritmo ha clasificado correctamente. Por tanto, miden la cantidad de ejemplos normales que detecta el método [4].

Se calcula mediante la siguiente ecuación, donde utilizamos los valores de la matriz de confusión:

$$TNR = \frac{TP}{TP + FP}$$

4.4 Precisión

En inglés conocido como (Precision). Se refiere a lo cerca que está el resultado de una medición del valor verdadero. Se representa por la proporción entre los positivos reales predichos por el algoritmo y todos los casos positivos [4].

En forma práctica es el porcentaje de casos positivos detectados. Se calcula a partir de la siguiente formula y de los datos de la matriz de confusión:

$$P = \frac{TP}{FP + TP}$$

4.5 ROC-AUC

Una curva ROC (curva de característica operativa del receptor) es una gráfica que muestra el rendimiento de los modelos en todos los umbrales de clasificación [4].

ROC es una curva de probabilidad y AUC (área bajo la curva) representa el grado o medida de separabilidad. Será la métrica AUC, que representa un valor del área que queda por debajo de la curva ROC, la encargada de comparar unos modelos con otros indicando cuánto es capaz el modelo de distinguir entre clases. Cuanto más alto es el AUC, mejor es el modelo para predecir las anomalías correctamente y las clases normales también.

La curva ROC se traza con la sensibilidad frente la especificidad, donde la sensibilidad está en el eje Y, y el eje X está compuesto por 1- especificidad.

La evaluación de esta medida se clasifica de la siguiente forma:

- Valor cerca de 1: El modelo es excelente, tiene una gran capacidad de separabilidad. En este caso es perfectamente capaz de distinguir entre la clase positiva y la clase negativa.

- Valor cerca de 0.5: El modelo no tiene ninguna capacidad de separación de clases. Es la peor situación y el modelo no tiene capacidad de discriminación para distinguir entre clase positiva y negativa.

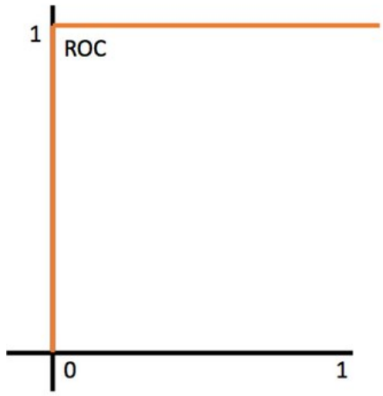

Figura 1. Situación ideal (AUC = 1)

Cuando dos curvas no se superponen en absoluto, el modelo tiene una medida ideal de separación. Es perfectamente capaz de distinguir entre clase positiva y clase negativa (figura 1).

Cuando dos distribuciones se superponen, introducimos errores. Dependiendo del umbral, podemos minimizarlos o maximizarlos. Cuando AUC es 0.7, significa que hay 70% de probabilidad de que el modelo pueda distinguir entre clase positiva y clase negativa (figura 2).

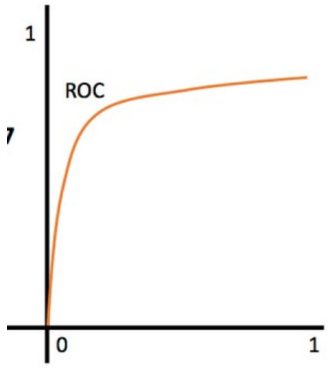

Figura 2. Existe superposición (AUC = 0.7)

La peor situación se produce cuando el AUC es aproximadamente 0.5, el modelo no tiene capacidad de discriminación para distinguir entre clase positiva y clase negativa (figura 3).

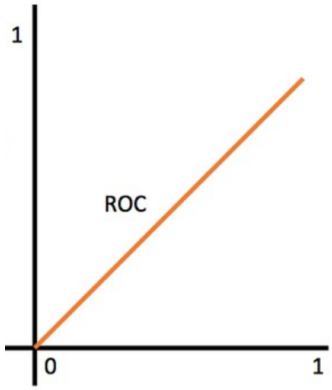

Figura 3. Peor situación (AUC = 0.5)

Cuando AUC es aproximadamente 0, el modelo en realidad está correspondiendo las clases. Significa que el modelo predice la clase negativa como una clase positiva y viceversa.

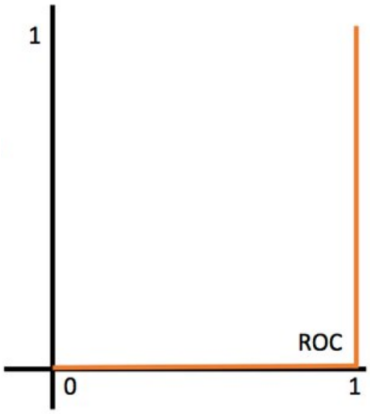

Figura 4. Clases al revés (AUC = 0)

Referencias

[1] C.C. Aggarwal. “Outlier analysis second edition”. Springer International Publishing, 2º edicion, 465 páginas. 2016.

[2] V. Chandola, A. Banerjee, V. Kumar. Anomaly detection: A survey. ACM computing surveys (CSUR), 41(3), 1-58. 2009.

[3] Narkhede S. Understanding AUC-ROC Curve. Medium. Towards Data Science; 2018. Disponible en: https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5

[4] Matriz de confusión. Disponible en: https://www.interactivechaos.com/manual/tutorial-de-machine-learning/matriz-de-confusion