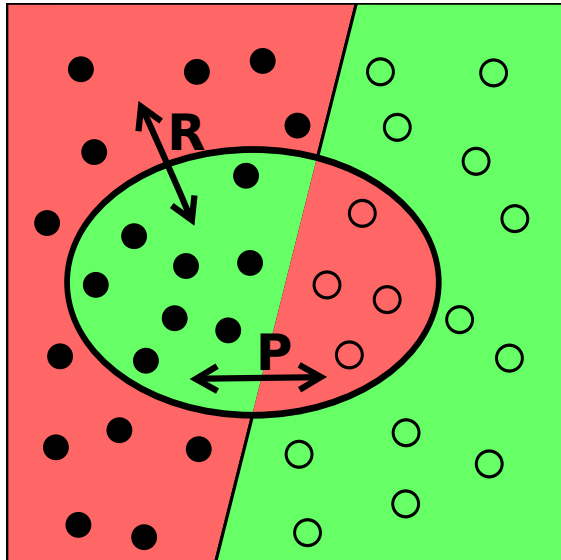


Precisión y exhaustividad



En la figura los ítems relevantes se encuentran en espacio ubicado a la izquierda de la línea recta, mientras que los ítems recuperados se encuentran en la zona oval. Las zonas de color rojo representan errores. A la izquierda se encuentran los ítems relevantes que no han podido ser recuperados (falsos negativos), mientras que a la derecha se encuentran aquellos recuperados que no son relevantes (falsos positivos).

La **precisión y exhaustividad** (denominado a veces como **exhaustividad y precisión**) es una métrica empleada en la medida del rendimiento de los sistemas de **búsqueda y recuperación de información y reconocimiento de patrones**. En este contexto se denomina precisión (denominado igualmente valor positivo predicho) como a la fracción de instancias recuperadas que son relevantes, mientras **recall** (denominado igualmente **sensibilidad o exhaustividad**) es la fracción de instancias relevantes que han sido recuperadas.^[1] Tanto la precisión como la exhaustividad son entendidas como medidas de la relevancia. Para entender mejor el concepto, supongamos de la existencia de un programa que reconoce perros en fotografías, dicho programa reconoce 7 perros en una escena que contiene 9 perros y algunos gatos. Si 4 de las identificaciones han sido correctas, pero 3 eran gatos, el programa tendrá una precisión de 4/7 mientras que posee una sensibilidad de 4/9. Otro ejemplo en el que participa un motor de búsqueda que, ante una consulta dada, retorna 30 páginas de las cuales sólo 20 son relevantes dejando 40 páginas relevantes fuera de la búsqueda. Este motor tendrá entonces una precisión de $20/30 = 2/3$ mientras que su sensibilidad es $20/60 = 1/3$.

Para un usuario la situación ideal es aquella en la que exis-

te una precisión y exhaustividad alta (es decir muy cercana a 1). A esta situación se la denomina *utilidad teórica*. Con el objeto de ponderar y ver cual lejano se encuentran ambas medidas de la utilidad teórica, suele emplearse los valores de ambas métricas combinadas en una media armónica denominada **valor-F**.

1 Concepto

1.1 Precisión

El concepto de precisión fue propuesto inicialmente por Salton en 1983.^[2] Algunos autores lo emplearon como un factor de pertinencia.^[3] La precisión es el ratio entre el número de documentos relevantes recuperados entre el número de documentos recuperados. Acorde con la definición se tiene la siguiente expresión:

$$\text{Precisión} = \frac{|\{\text{documentos relevantes}\} \cap \{\text{documentos recuperados}\}|}{|\{\text{documentos recuperados}\}|}$$

De esta forma, cuanto más se acerque el valor de la precisión al valor nulo, mayor será el número de documentos recuperados que no consideren relevantes. Si por el contrario, el valor de la precisión es igual a uno, se entenderá que todos los documentos recuperados son relevantes. Esta forma de entender la precisión introduce el concepto de ruido informativo y de silencio informativo.

1.2 Exhaustividad

La exhaustividad se emplea en menor medida que la precisión. Algunos autores suelen definirlo en lengua castellana como “rellamada” procedente del término inglés “recall”, en otros casos como “recobrado”. Este ratio viene a expresar la proporción de documentos relevantes recuperados, comparado con el total de los documentos que son relevantes existentes en la base de datos, con total independencia de que éstos, se recuperen o no.^[4] Se definió por primera vez a mediados del siglo XX.^[3] La ecuación en este caso se expresa como:

$$\text{Exhaustividad} = \frac{|\{\text{documentos relevantes}\} \cap \{\text{documentos recuperados}\}|}{|\{\text{documentos relevantes}\}|}$$

Si el resultado de esta fórmula arroja como valor 1, se tendrá la exhaustividad máxima posible, y esto viene a

indicar que se ha encontrado todo documento relevante que residía en la base de datos, por lo tanto no se tendrá ni ruido, ni silencio informativo: siendo la recuperación de documentos entendida como perfecta. Por el contrario en el caso que el valor de la exhaustividad sea igual a cero, se tiene que los documentos obtenidos no poseen relevancia alguna.

2 Referencias

- [1] Raquel Gómez Díaz. (2003), *La evaluación en recuperación de la información en línea*. “Hipertext.net”, núm. 1
- [2] Salton, G. y M. J. McGill., (1983), *Introduction to Modern Information Retrieval*. New York: McGraw Hill
- [3] Kent A. Et al., (1955), *Machine literature searching*. VIII. Operational Criteria for Designing Information Retrieval Systems American Documentation Abril, 6 (2) p. 93-101
- [4] Swets, J. A., (1963), *Information retrieval Systems Science*, 141 (3577): July 1963 p. 245-250

3 Véase también

- Sensibilidad y especificidad
- Valor-F

4 Origen del texto y las imágenes, colaboradores y licencias

4.1 Texto

- **Precisión y exhaustividad** *Fuente:* https://es.wikipedia.org/wiki/Precisi%C3%B3n_y_exhaustividad?oldid=92396916 *Colaboradores:* Benjavaleto, Tamorlan, NudoMarinero, Langtoolbot y Anónimos: 2

4.2 Imágenes

- **Archivo:Recall-precision.svg** *Fuente:* <https://upload.wikimedia.org/wikipedia/commons/e/e8/Recall-precision.svg> *Licencia:* Public domain *Colaboradores:* Trabajo propio *Artista original:* Nichtich

4.3 Licencia del contenido

- Creative Commons Attribution-Share Alike 3.0