

Solemne 3- Sistemas Inteligentes

martes 1 de diciembre de 2015

Profesor: Alejandro Figueroa

Ayudantes: Alexander Espina - Daniel Palomera

- Está prohibido el uso de teléfonos celulares durante el desarrollo de la prueba.
- La prueba debe responderse con un lápiz de tinta indeleble, de lo contrario no hay opción a correcciones.
- Cualquier alumno que sea sorprendido intentando copiar será calificado con una nota 1.
- Está prohibido conversar durante la prueba. Recuerde que su compañero puede estar concentrado y el ruido puede perturbarlo en el desarrollo de su prueba.
- Utilice sólo las hojas entregadas para escribir sus respuestas.

Pregunta 1 (20 puntos)

Considere que se tienen dos sistemas S_1 y S_2 , para los cuales se obtuvieron las siguientes métricas de desempeño:

	Precision	Recall	MRR	Accuracy
S_1	0.5	0.6	0.55	0.45
S_2	0.6	0.5	0.4	0.59

Se le pide calcular el F(B)-Score para S_1 y S_2 considerando los valores de $B=0, 1, 2, 3, 4$ y 5 . De acuerdo a F(B) ¿Qué sistema es mejor? Nótese que hablamos de una tarea de clasificación binaria de documentos.

Desarrollo:

B	1	2	3	4	5
S_1	0.5455	0.5769	0.5882	0.593	0.5954
S_2	0.5455	0.5172	0.5085	0.505	0.5032

En términos estándar, se utiliza $B=1$, y en ese sentido, ambos sistemas son iguales. Sin embargo, dependiendo de la tarea donde esté el clasificador va a ser mejor S_1 o S_2 . Hay aplicaciones que requieren un mayor Recall otras Precisión.

Pregunta 2 (30 puntos)

Se define Mean Average Precision (MAP) como sigue: El primer paso para determinar el MAP de un sistema consiste en calcular $P@k$ para cada una de las k posiciones de las salidas a evaluar. El segundo paso, consiste en calcular el Average Precisión (AP) para una de las salidas. Con este objetivo, utilizamos la siguiente fórmula:

$$AP = \frac{\sum_{k=1}^m P@k * I(k)}{m_1}$$

Donde $I(k)$ indica con un 1 ó 0 si hay o no un documento relevante en la posición “ k ”, m es la cantidad de documentos en la salida, y m_1 es número de documentos relevantes. El MAP se define como el promedio de los AP obtenidos para cada una de las salidas. Calcule el MAP para un sistema que produce la siguiente salida:

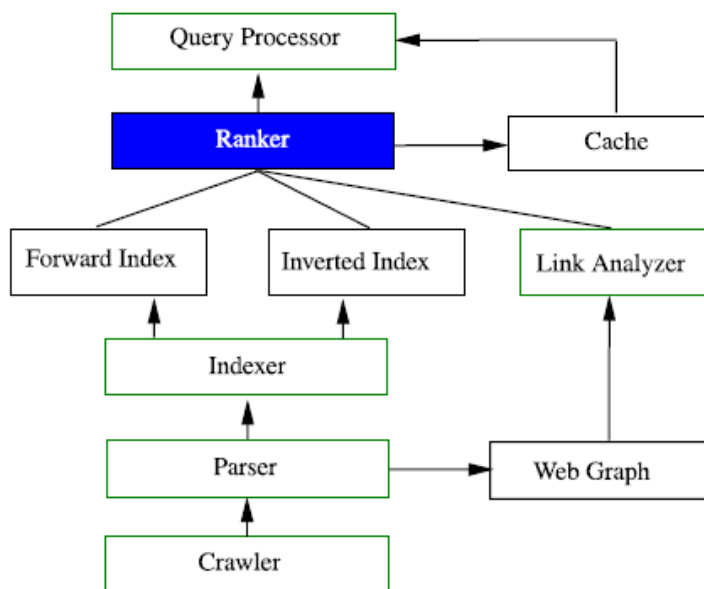
Posición	Salida 1	Salida 2
1	1	1
2	1	0
3	0	1
4	0	0
5	1	1

Desarrollo:

Posición	Salida 1	Salida 2	$P@k$ S1	$P@k$ S2	$P@k$ S1 * $I(k)$ S1	$P@k$ S2 * $I(k)$ S2	
1	1	1	1	1	1	1	
2	1	0	1	0.5	1	0	
3	0	1	0.666667	0.666667	0	0.6666666667	
4	0	0	0.5	0.5	0	0	
5	1	1	0.6	0.6	0.6	0.6	MAP
			Average ▶		0.8666666667	0.7555555556	0.8111111111

Pregunta 3 (10 puntos)

Esquematice la arquitectura típica de un buscador web.



Pregunta 4 (20 puntos)

Calcule el indicador AUC para un clasificador que obtiene los siguientes valores de TPR y FPR:

TPR	0.0	0.7	0.7	0.8	0.8	1.0	1.0	
FPR	0.0	0.0	0.3	0.3	0.5	0.5	1.0	

Desarrollo:

Para calcular el área bajo la curva, utilizamos la fórmula del área de un rectángulo. Entonces, el área bajo la curva está dada por $0,7 \cdot 0,3 + 0,8 \cdot (0,5 - 0,3) + 1,0 \cdot (1,0 - 0,5) = 0,21 + 0,8 \cdot 0,2 + 1,0 \cdot 0,5 = 0,21 + 0,16 + 0,5 = 0,87$.

Pregunta 5 (20 puntos)

Para la siguiente pregunta considere que se tiene:

- Se utiliza un algoritmo de clustering, más precisamente k-means.
- El valor del parámetro “k” es dos.
- Un algoritmo genético es quien busca la mejor asignación de cada dato a cada clúster.
- Asuma que no se tienen las etiquetas.

¿Cuál sería la función objetivo de este algoritmo genético? Detalle claramente.

Analice además qué podría hacerse si se tiene las etiquetas.

Desarrollo:

Se puede utilizar varias funciones objetivos, por ejemplo la minimización del error cuadrático, la varianza interna, se puede maximizar el índice de Dunn. Si se tienen las etiquetas, se pueden buscar los centroides que maximizan el purity o minimizan la entropía.