

AVALIAÇÃO #2

1. **(2.0)** Como podemos representar as preferências de um usuário para recomendação personalizada baseada em conteúdo?
2. **(2.0)** Para a questão anterior, como poderíamos representar usuários novos (*cold-start users*)?
3. **(2.0)** Learning to rank é uma técnica tipicamente utilizada para reordenação (*reranking*) dos top  $k$  itens retornados por algum modelo de recomendação mais simples, em vez de diretamente sobre todos os itens disponíveis no sistema. Que fatores devem ser considerados na escolha de um valor ótimo de  $k$ ?
4. **(5.0)** Considere as seguintes matrizes (a e b) e fórmulas de referência (c):

	$i_0$	$i_1$	$i_2$	$i_3$	$i_4$	$i_5$
$u_0$		4		3		3
$u_1$			1		3	4
$u_2$	1	4			2	
$u_3$	2	3	3			
$u_4$	4			4		2
$u_5$	5		1			5
$u_6$			1		2	4
$u_7$			5	1	3	
$u_8$	2			1	4	
$u_9$		3	5	4		

(a) matriz de avaliações

	$i_0$	$i_1$	$i_2$	$i_3$	$i_4$	$i_5$
$t_0$	5	7	2	4	0	0
$t_1$	0	0	0	0	8	3
$t_2$	7	9	7	0	0	7
$t_3$	6	0	4	2	0	0
$t_4$	9	5	0	5	6	8
$t_5$	0	9	3	4	0	6

(b) frequências (TF)

**Predição (Rocchio) para o usuário  $u$  e o item  $i$ :**

- $\vec{u}$ : centróide do usuário,  $\vec{u} = \frac{1}{|I_u|} \sum_{j \in I_u} r_{uj} \vec{j}$ 
  - $I_u$ : itens avaliados por  $u$
  - $r_{uj}$ : avaliação do usuário  $u$  sobre o item  $j$
- $s_{\vec{u}\vec{i}}$ : similaridade entre o usuário  $u$  e o item  $i$

**Similaridade (cosseno) entre os vetores  $\vec{u}$  e  $\vec{i}$ :**

$$s_{\vec{u}\vec{i}} = \frac{\sum_{p=1}^d u_p i_p}{\sqrt{\sum_{p=1}^d u_p^2} \sqrt{\sum_{p=1}^d i_p^2}}$$

- $u_p$ :  $p$ -ésima dimensão do vetor  $\vec{u}$
- $i_p$ :  $p$ -ésima dimensão do vetor  $\vec{i}$
- $d$ : número de dimensões dos vetores

(c) fórmulas de referência

Para um dado usuário-alvo  $u_x$ , onde  $x$  é o último dígito de seu número de matrícula (e.g., se seu número de matrícula é 2021081014, seu usuário-alvo é  $u_4$ ), **recomende itens previamente não-avaliados por  $u_x$  usando o algoritmo Rocchio para recomendação baseada em conteúdo, com similaridade do cosseno, e vetores baseados em frequência (TF)**. Para cada item recomendado, indique claramente (1) a predição computada e (2) os cálculos intermediários que levaram a essa predição. A matriz de avaliações mostrada acima **não** deve ser normalizada.

5. **(4.0)** Considere (a) as restrições  $\{c_0, \dots, c_3\}$  do usuário e (b) o catálogo de itens  $\{i_0, \dots, i_4\}$  a seguir, ambos expressos em termos de cinco atributos dos itens catalogados, a saber, atributos  $A, B, C, D, E$ :

$$\begin{aligned} c_0 &: (C \neq 1) \wedge (A > 40) \\ c_1 &: (B < 150) \\ c_2 &: (D < 15) \\ c_3 &: (D < 50) \wedge (E > 49) \end{aligned}$$

(a) restrições do usuário

	$A$	$B$	$C$	$D$	$E$
$i_0$	100	199	21	15	20
$i_1$	15	15	1	80	30
$i_2$	72	300	30	18	40
$i_3$	33	9000	40	5	50
$i_4$	10	13	50	35	60

(b) catálogo de itens

Para um recomendador baseado em restrições (*constraints*), liste:

- (a) Os conjuntos conflituosos de restrições (incluindo as restrições em cada conjunto).
- (b) Os diagnósticos (incluindo as restrições em cada diagnóstico) para os conjuntos conflituosos encontrados.