

Machine Learning

Filtrage Collaboratif item-based

Filtrage Collaboratif item-based

	item 1	item 2	item 3	item 4
u1	1	2	5	?
u2	5	1	4	3
u3	?	?	1	5
u4	3	3	?	4

Pour remplir les cellules manquantes, au lieu de traiter les lignes on traite les colonnes.

Une première solution est de transposer la matrice et d'appliquer l'algorithme décrit pour le user-based.

Filtrage Collaboratif item-based

Méthode couramment utilisée \Rightarrow régression simplifiée.

	item 1	item 2	item 3	item 4
u1	1	2	5	?
u2	5	1	4	3
u3	?	?	1	5
u4	3	3	?	4

Filtrage Collaboratif item-based

Sachant la colonne item1, on cherche uniquement une valeur de biais telle que :

$$item4 = item1 + b_1$$

Pour ce faire, on prend le biais moyen entre ces deux colonnes :

$$b_1 = \frac{\sum_{i \in I} (item4_i - item1_i)}{\#I}, \text{ dans notre exemple : } b_1 = \frac{(3-5)+(4-3)}{2} = -0.5$$

	item 1	item 2	item 3	item 4
u1	1	2	5	?
u2	5	1	4	3
u3	?	?	1	5
u4	3	3	?	4

Filtrage Collaboratif item-based

On calcul alors les biais pour chaque colonne, nous donnant une prediction de la valeur $u1(item4)$ sachant $u1(item1)$, une autre sachant $u1(item2)$ et une dernière sachant $u1(item3)$.

La prédiction finale pour $u1(item4)$ est la moyenne pondérée de toutes ces prédictions.

$$b_1 = -0.5 \Rightarrow u1(item4)_1 = 0.5$$

$$b_2 = 1.5 \Rightarrow u1(item4)_2 = 2.5$$

$$b_3 = 1.5 \Rightarrow u1(item4)_3 = 2.5$$

$$u1(item4) = \frac{\#I_1 * 0.5 + \#I_2 * 2.5 + \#I_3 * 2.5}{\#I_1 + \#I_2 + \#I_3} = \frac{11}{6} \approx 1.83$$

	item 1	item 2	item 3	item 4
u1	1	2	5	1.83
u2	5	1	4	3
u3	?	?	1	5
u4	3	3	?	4

- Calculs simples à mettre en oeuvre et à tenir à jour
- Beaucoup de paramètres à maintenir quand on a beaucoup d'item :
 $\left(\frac{\#item(\#item-1)}{2} \right)$
- "Cold start problem"