

Machine Learning

Filtrage Collaboratif - Aujourd'hui

Filtrage Collaboratif - SVD

Utilisation de SVD pour réduire le nombre de dimensions du problème.

$M = U^P * V^P * I^P$, où :

M est la matrice user/item de \mathbf{R}^{n*m}

U^P la matrice des vecteurs de base orthonormée de \mathbf{R}^n (input)

V^P la matrice diagonale des valeurs singulières de U

I^P la matrice des vecteurs de base orthonormée de \mathbf{R}^m (output)

$$M = \begin{array}{|c|c|c|} \hline 2 & 5 & ? \\ \hline 1 & 4 & 3 \\ \hline ? & 1 & 5 \\ \hline 3 & ? & 4 \\ \hline \end{array} =$$

$$\begin{array}{|c|c|} \hline U_{(1,1)}^P & U_{(1,2)}^P \\ \hline U_{(2,1)}^P & U_{(2,2)}^P \\ \hline U_{(3,1)}^P & U_{(3,2)}^P \\ \hline U_{(4,1)}^P & U_{(4,2)}^P \\ \hline \end{array} * \begin{array}{|c|c|} \hline V_{(1)}^P & 0 \\ \hline 0 & V_{(2)}^P \\ \hline \end{array} * \begin{array}{|c|c|c|} \hline I_{(1,1)}^P & I_{(1,2)}^P & I_{(1,3)}^P \\ \hline I_{(2,1)}^P & I_{(2,2)}^P & I_{(2,3)}^P \\ \hline \end{array}$$

Avantages :

On choisit le nombre de valeurs singulières de notre transformation (PCA).

Filtrage collaboratif user-based $\Rightarrow U^p$

Filtrage collaboratif item-based $\Rightarrow I^p$

La SVD est extrêmement coûteuse sur des grandes matrices :

$$O(\min(m * n^2, m^2 * n))$$

Heureusement des algorithmes approchés existent. Ces approches tirent parti du fait que les matrices sont généralement sparses et que l'on ne s'intéresse qu'à un nombre limité de valeurs singulières.

Par exemple, on peut calculer les 20 premières valeurs singulières d'une matrice 100k x 100k avec 1M valeurs non-nulles en moins d'une seconde (redsvd).

Filtrage Collaboratif - Embeddings

Utilisation d'Embeddings sur les utilisateurs et les items afin de pouvoir exprimer la matrice des ratings ainsi :

$$M = U^{embed} * I^{embed}$$

Pour Faire une recommandation :

$$\operatorname{argmax}(u_{query} * I^t)$$

Jusqu'à présent, la nature des items était sans importance.

Que les items soient des livres ou des voitures, l'algorithme ne faisait aucune différence.

On peut utiliser les détails textuels, les images (, etc...) sur les items afin d'améliorer les recommandations. (Embeddings, Catégorisation des items)