

Matrices creuses et valeurs propres

1. Quelle est la définition d'une matrice creuse ? D'un vecteur creux ?
2. Imaginez que vous ayez une matrice creuse $n \times m$ sur \mathbb{R} . Quelle est la complexité de calcul de la multiplication de cette matrice par un vecteur quelconque ?
3. Décrivez la méthode de la puissance itérée pour le calcul de valeurs propres ? Comment fait-on pour extraire la seconde valeur propre ?
4. En utilisant la méthode de la puissance itérée, trouvez le premier vecteur propre de la matrice suivante :

$$\begin{pmatrix} 3 & 0 & 1 \\ 0 & 2 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

Pourquoi est-on sûr que les valeurs propres de cette matrice sont réelles et non complexes ?

5. Expliquez pourquoi Google utilise la méthode de la puissance itérée à la place d'une autre méthode d'extraction de vecteur/valeur propre ?

HITS

1. Sous quelles conditions l'algorithme HITS converge-t-il ?
2. Concernant l'algorithme de construction du graphe de base BaseSubGraph, que se passe-t-il si d est trop grand ? Et s'il est trop petit ?
3. Supposons que d soit inférieur à $|P|$ (initialement). Combien faut-il tirer au hasard de sous-ensembles de P de taille d pour être sûr à 95% que l'un d'eux fait partie des 5% meilleurs sous-ensembles (on fait l'hypothèse que tous les sous-ensemble de taille d ont la même chance d'être le meilleur) ?

PageRank

1. Proposer une méthode permettant d'augmenter artificiellement la valeur PageRank d'une page Web.
2. Proposer une méthode inspirée de PageRank qui permet de ne pas attribuer un Page rank élevé aux pages qui sont des pages de spam. Comparer votre algorithme à l'algorithme TrustRank (<http://www.vldb.org/conf/2004/RS15P3.PDF>).