

Calcul MapReduce d'une matrice de liens dans un modèle PageRank

Exercice 1

Auteur : Ghada Zeineb Malek

Date : December 17, 2024

Contents

1	Introduction	2
2	Représentation de la matrice M	2
3	Calcul de la norme des vecteurs lignes	2
3.1	Étape Map	2
3.2	Étape Reduce	2
3.3	Exemple : Calcul de la norme	3
4	Calcul du produit matrice-vecteur $M \cdot W$	3
4.1	Étape Map	3
4.2	Étape Reduce	3
4.3	Exemple : Calcul du produit matrice-vecteur	3
5	Conclusion	3

1 Introduction

Dans cet exercice, nous considérons une matrice M de dimension $N \times N$, représentant des liens entre un très grand nombre de pages web. Chaque élément M_{ij} représente un lien entre la page P_i et la page P_j , étiqueté par un poids positif reflétant son importance.

L'objectif est de proposer un modèle pour structurer cette matrice, en s'inspirant du cas *PageRank* utilisé par Google. La collection C ainsi obtenue sera utilisée pour :

1. Calculer la norme des vecteurs lignes de la matrice.
2. Calculer le produit de la matrice M avec un vecteur W .

Le traitement sera spécifié sous forme d'algorithmes utilisant *MapReduce*, qui est adapté aux traitements distribués pour les données de grande taille.

2 Représentation de la matrice M

Chaque ligne i de la matrice M peut être vue comme un vecteur $V_i = (M_{i1}, M_{i2}, \dots, M_{iN})$. La matrice sera représentée sous forme de documents structurés dans une collection C telle que :

- Chaque document correspond à une ligne de la matrice.
- Un document D_i représente le vecteur associé à la page P_i : $D_i = (M_{i1}, M_{i2}, \dots, M_{iN})$.
- Le poids M_{ij} représente la contribution de la page P_j à la page P_i .

Ainsi, la collection C est composée de documents $\{D_1, D_2, \dots, D_N\}$.

3 Calcul de la norme des vecteurs lignes

La norme d'un vecteur $V_i = (v_1, v_2, \dots, v_N)$ est définie par :

$$\|V_i\| = \sqrt{v_1^2 + v_2^2 + \dots + v_N^2}.$$

Nous allons maintenant détailler le calcul de cette norme en utilisant le modèle MapReduce.

3.1 Étape Map

- **Entrée** : Un document D_i contenant le vecteur $(M_{i1}, M_{i2}, \dots, M_{iN})$.
- **Opération** : Chaque valeur M_{ij} est élevée au carré, puis émise avec la clé i .
- **Sortie** : Des paires clé-valeur (i, M_{ij}^2) pour chaque j .

3.2 Étape Reduce

- **Entrée** : Toutes les valeurs (M_{ij}^2) pour une clé donnée i .
- **Opération** : La somme des carrés est calculée : $\text{sum} = \sum_{j=1}^N M_{ij}^2$.
- **Sortie** : La norme est calculée comme $\|V_i\| = \sqrt{\text{sum}}$.

3.3 Exemple : Calcul de la norme

Soit la matrice $M = \begin{bmatrix} 4 & 5 & 6 \\ 1 & 2 & 3 \end{bmatrix}$.

- Pour la ligne $V_1 = (4, 5, 6)$, la norme est :

$$\|V_1\| = \sqrt{4^2 + 5^2 + 6^2} = \sqrt{16 + 25 + 36} = \sqrt{77}.$$

- Pour la ligne $V_2 = (1, 2, 3)$, la norme est :

$$\|V_2\| = \sqrt{1^2 + 2^2 + 3^2} = \sqrt{1 + 4 + 9} = \sqrt{14}.$$

4 Calcul du produit matrice-vecteur $M \cdot W$

Nous voulons calculer le produit de la matrice M avec un vecteur $W = (w_1, w_2, \dots, w_N)$. Le résultat est un vecteur $\phi = (\phi_1, \phi_2, \dots, \phi_N)$, où chaque élément ϕ_i est donné par :

$$\phi_i = \sum_{j=1}^N M_{ij} w_j.$$

Nous détaillons maintenant l'algorithme MapReduce pour le calcul de ce produit.

4.1 Étape Map

- **Entrée** : Un document D_i contenant le vecteur $(M_{i1}, M_{i2}, \dots, M_{iN})$.
- **Opération** : Pour chaque j , on calcule le produit $M_{ij} \cdot w_j$.
- **Sortie** : Des paires clé-valeur $(i, M_{ij} \cdot w_j)$.

4.2 Étape Reduce

- **Entrée** : Toutes les valeurs $(M_{ij} \cdot w_j)$ pour une clé donnée i .
- **Opération** : La somme des produits est calculée : $\phi_i = \sum_{j=1}^N M_{ij} \cdot w_j$.
- **Sortie** : (i, ϕ_i) , où ϕ_i est le produit pour la ligne i .

4.3 Exemple : Calcul du produit matrice-vecteur

Soit la matrice $M = \begin{bmatrix} 4 & 5 & 6 \\ 1 & 2 & 3 \end{bmatrix}$ et le vecteur $W = (2, 1, 3)$.

- Pour la ligne $V_1 = (4, 5, 6)$, le produit est :

$$\phi_1 = 4 \cdot 2 + 5 \cdot 1 + 6 \cdot 3 = 8 + 5 + 18 = 31.$$

- Pour la ligne $V_2 = (1, 2, 3)$, le produit est :

$$\phi_2 = 1 \cdot 2 + 2 \cdot 1 + 3 \cdot 3 = 2 + 2 + 9 = 13.$$

Ainsi, le produit $M \cdot W$ est $\phi = (31, 13)$.

5 Conclusion

Ce document a spécifié deux traitements MapReduce pour :

- Calculer la norme des vecteurs lignes de la matrice M .
- Calculer le produit de la matrice M avec un vecteur W .

Ces traitements sont adaptés aux données massives et tirent parti de la parallélisation offerte par MapReduce.