

Contexte

- Ensemble d'utilisateurs
- Caractéristiques utilisateurs observées
- Potentiel de calcul des utilisateurs

Objectifs

- Regrouper les utilisateurs en tribus
- Répartir le calcul

Contraintes

- Informations privées
- Limitation de la communication

1 Modélisation

2 Résolution

Notations

- Réseau des utilisateurs $G = (V, E)$, $V = \{1, \dots, N\}$
- Caractéristiques $\mathbf{X} = (X_i)_{1 \leq i \leq N}$
ex : SMS, a écouté [XXX], a partagé [YYY]
- Distance entre utilisateurs :

$$D : \begin{cases} V \times V & \rightarrow \mathbb{R}_+ \\ (x, y) & \mapsto D(x, y) \end{cases}$$

$$D(x, y) = 0 \iff x = y$$

$$D(x, y) = \|x - y\|_p$$

Formulation

- Objectif : partitionner en tribus, par affinités
- Problème à résoudre : trouver une partition \mathcal{P}^* solution de

$$\min_{\mathcal{P}} W(\mathcal{P}) = \frac{2}{N(N-1)} \sum_{1 \leq i < j \leq N} D(X_i, X_j) \Phi_{\mathcal{P}}(i, j),$$

$$\Phi_{\mathcal{P}}(i, j) = \begin{cases} 1 & \text{si } i \text{ et } j \text{ dans le même cluster} \\ 0 & \text{sinon} \end{cases}$$

1 Modélisation

2 Résolution

Contraintes

- Contraintes supplémentaires :
 - Distribution du calcul
 - Limitation du transfert de données
- Problème d'optimisation distribuée

Énoncé

Nouvelle formulation du problème :

$$\min_{\mathcal{P}} W(\mathcal{P}) = \frac{1}{N} \sum_{1 \leq i \leq N} f_i(\mathbf{x}),$$

où

$$f_i(\mathbf{x}) = \frac{1}{N} \sum_{1 \leq j \leq N} D(X_i, X_j) \Phi_{\mathcal{P}}(i, j)$$

- Principe
 - Utilisateur i effectue seul le calcul de f_i
 - Communication restreinte : voisinage dans G , taux de données
- Intérêt
 - Problème décomposé suivant chaque utilisateur
 - Nombreuses méthodes si problème convexe
 - Possibilité d'optimiser en ligne
- Challenges
 - f_i non calculable par i
 - Estimation de f_i