

Travail individuel en Fouille de données

L'objectif dans ce travail individuel est de savoir appliquer le clustering pour l'amélioration de la synchronisation entre des appareils électroniques en minimisant le temps de synchronisation et en minimisant les interférences entre les appareils.

1. Description du cas :

Nous disposons d'un ensemble de 05 appareils électroniques placés de manière fixe dans une pièce carrée de 2500 m² (50X50). Chaque appareil est décrit par deux coordonnées : l'abscisse et l'ordonnée par rapport à un coin de la pièce. Les appareils doivent fonctionner ensemble pour réaliser des tâches. Pour cela, ils doivent d'abord être synchronisés en échangeant des messages (flèches bidirectionnelles). La rapidité de la synchronisation dépend de la distance entre les appareils : deux appareils proches se synchronisent plus vite que deux appareils éloignés. Cependant, la distance entre les appareils influe négativement sur le degré d'interférence avec les autres signaux émis par chaque appareil: deux appareils proches ont un fort degré d'interférence et inversement.

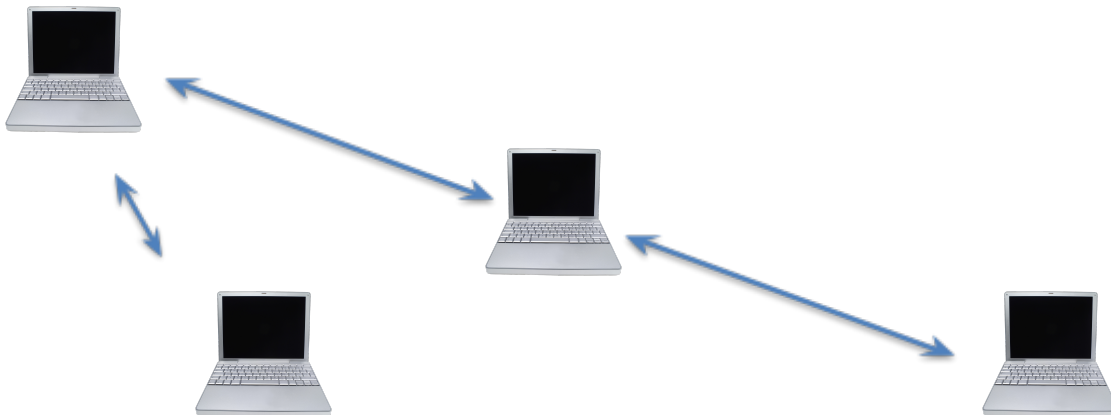


Figure: Disposition des appareils et synchronisation

2. Mode de fonctionnement

- *Regroupement des appareils* : on procède à deux modes de regroupements. Le premier mode est aléatoire : on affecte chaque appareil à un groupe. Par la suite, à chaque étape, on regroupe deux groupes aléatoirement, jusqu'à obtenir le dernier groupe qui contient tous les appareils. Le deuxième mode consiste à utiliser le clustering hiérarchique agglomératif en deux sous-modes : le linkage single et le linkage complet..
- *Calcul du temps d'envoi du signal et calcul du degré d'interférence* : on fait les suppositions suivantes : (1) La durée de synchronisation est fixe par mètre (ex : 2 sec/m). (2) une fois deux appareils ou plus regroupées, la synchronisation entre deux groupes d'appareils se fait

entre deux appareils, chacun d'un seul groupe. Les appareils qui communiquent dépendent du mode de regroupement (aléatoirement ou selon le type de linkage utilisé dans le clustering). Le calcul de la durée et du degré d'interférence se base sur les appareils qui communiquent. Les degrés d'interférences peuvent être déduits du tableau suivant :

Distance	[0, 5]]5, 20]]20, 35]]35, 50]]50, 65]]65, 71]
Degré d'interférence.	6	5	4	3	2	1

3. Optimisation de la synchronisation : on cherche à optimiser la synchronisation en terme de temps et de degré d'interférences. On teste le mode aléatoire, puis le clustering en deux sous-modes, dépendant du type le linkage. On compare et on propose le meilleur ordre de synchronisation.

Travail demandé : les étapes du travail sont comme suit

1. **Préparation des données transactionnelles :** répartir 05 appareils sur l'espace et fixer le temps de synchronisation par mètre.
2. **Calcul du temps de synchronisation et du degré d'interférence par organisation aléatoire:** choisir aléatoirement le regroupement étape par étape de deux groupes et en déduire la durée totale de synchronisation et le degré d'interférence.
3. **Etablissement de l'ordre de synchronisation par clustering hiérarchique :** définir et justifier le type de distance entre les groupes d'appareils puis utiliser le clustering hiérarchique agglomératif en deux modes de linkage pour définir l'ordre de synchronisation des groupes d'appareils.
4. **Calcul des nouveaux temps et degrés :** à partir de l'étape précédente, en déduire le temps global de synchronisation (somme des temps) et le degré d'interférence global (somme des degrés) et ce, pour chaque mode de linkage
5. **Comparaison des temps de synchronisation et des degrés d'interférence :** comparer les résultats de chaque solution. Donner et justifier la meilleure solution.
6. **Question 1 :** Supposons que les appareils sont en mouvement dans l'espace. Quel est l'effet de ce mouvement sur le clustering, i.e. que faut-il changer dans les clusterings précédents? (il n'est pas demandé de refaire le travail)
7. **Question 2:** le linkage par centroïde est-il applicable? justifier la réponse.

Modalité d'envoi du travail

- Le travail est **individuel**
- Le travail est à rédiger comme un rapport sous forme d'un document word (ou autre) mais doit être obligatoirement converti **en format PDF**.
- Le rapport doit contenir une page de garde mentionnant le nom, prénom et spécialité de l'étudiant(e), les autres pages contiendront les étapes et questions du travail.
- Le rapport en format PDF est à envoyer à l'adresse email d_boukraa@esi.dz au plus tard le **jeudi 10 septembre 2020 à 23h59**.
- Toute question relative au travail doit être adressée à l'adresse email ci-dessus.