



Projet

Métaheuristique

Réalisées par

Briere Assia

El ghayatii Loubna

Ghommat Khadija



Problème d'optimisation

L'objectif est de trouver une permutation optimale des mots dans chaque phrase donnée, de manière à réduire au minimum la perplexité mesurée par un modèle de langage pré-entraîné (Gemma 2 9B). La perplexité reflète la capacité du modèle à prédire une séquence fluide et cohérente.

les variables de décision

représentent les positions des mots dans une séquence donnée

Pour une séquence donnée de n mots, les variables de décision sont :

$$x_i \in \{1, 2, \dots, n\}, \forall i \in \{1, 2, \dots, n\}$$

i représente l'indice du mot initial dans la liste fournie.

x_i représente la position du mot i dans la permutation finale.

x_i indique l'ordre des mots dans la séquence réarrangée.

Séquence initiale :

["advent", "chimney", "elf", "family", "fireplace", "gingerbread", "mistletoe", "ornament", "reindeer", "scrooge"]

Les mots sont indexés comme suit :

{1:"advent", 2:"chimney", 3:"elf", ..., 10:"scrooge"}

Les variables de décision x_1, x_2, \dots, x_{10} assignent une position à chaque mot dans l'ordre final.

Fonction objectif :

Minimiser la perplexité d'une séquence de mots réarrangée

Soit $P(x)$ la perplexité calculée par le modèle pour une permutation x des mots de la séquence.

La fonction objective est : **$\min P(x)$**

La perplexité $P(x)$ est calculée par le modèle de langage ("Gemma 2 9B").

Contraintes

- Chaque séquence réarrangée doit contenir exactement les mêmes mots que la séquence originale, mais dans un ordre différent. Les mots ne peuvent pas être ajoutés, supprimés ou modifiés.
- Chaque mot doit apparaître exactement une fois.

Espace des solutions

L'espace des solutions est constitué de toutes les permutations valides des mots de la séquence originale S .

pour résoudre ce problème d'optimisation à l'aide de métaheuristiques génétiques

1.Représentation des individus (solutions)

Chaque individu (solution) dans la population est une permutation des mots de la séquence donnée dans une ligne du dataset.

advent chimney elf family fireplace gingerbread mistletoe ornament reindeer scrooge

Un individu pourrait être :

elf chimney advent fireplace mistletoe ornament family reindeer scrooge gingerbread

2. Population initiale

Créez une population initiale de N individus, où chaque individu est une permutation aléatoire des mots de la séquence donnée.

- family scrooge elf fireplace ornament chimney gingerbread advent reindeer mistletoe
- reindeer chimney fireplace gingerbread elf mistletoe scrooge advent family ornament
- chimney reindeer advent elf mistletoe ornament family scrooge fireplace gingerbread

3. Fonction objectif

$$\underline{\text{Fitness}(S) = \text{Perplexity}(S)}$$

4. Sélection

5. Croisement

6. Mutation

7. utilisation de Gradient descent comme methode de local search