Rapport de projet

Aide à la Décision

Florian Posez & Théophile Romieu

M2 IAD
UFR Maths-Info
Université Paris-Cité
Novembre 2024

1 Introduction

1.1 Présentation du projet

L'énoncé de ce projet d'aide à la décision provient d'une situation réelle où des commerciaux de l'entreprise *Pfizer Turkey* sont assignés à des zones géoraphiques. Le but de ces commerciaux est de rendre visite aux médecins pour leur vendre les nouveaux médicaments produits par *Pfizer Turkey*, ils sont aussi chargés de répondre aux différentes questions des médecins quant aux potentiels effets secondaires et risques de ces médicaments.

Les zones d'activité des commerciaux sont appelées "briques", chaque brique possède un indicateur de charge de travail qui varie suivant la quantité de travail à effectuer pour rendre visite à tous les médecins se trouvant dans la brique. Dans la configuration actuelle des assignations, chaque commercial possède une brique centrale où se trouve son bureau. Le commercial n'a pas de distance à parcourir quand il doit travailler dans sa brique centrale, il doit cependant se déplacer depuis sa brique centrale jusqu'aux autres briques pour pouvoir effectuer son travail. Une brique est assignée à un seul commercial, les briques peuvent être réassignées à d'autres commerciaux pour des raisons d'optimisation, seules les briques centrales ne peuvent pas être réassignées.

Le but de ce projet est d'optimiser l'assignation des briques pour chaque commercial en se basant sur deux mesures : la distance totale parcourue par les commercieaux et le nombre de briques réassignées. Nous souhaitons minimiser ces mesures. De plus la charge de travail doit être répartie entre les différents commerciaux, cela se traduira par des contraintes sur le modèle. Il s'agit d'un problème d'optimisation multi-objectif qui sera défini dans la partie 1.2. Pour calculer ces mesures, nous avons besoin de trois matrices :

- D la matrice des distances où $D_{i,j}$ donne la distance entre la brique centrale du commercial j et la brique i.
- A la matrice des assignations où $A_{i,j} = 1$ si le commercial j est assigné à la brique i et $A_{i,j} = 0$ sinon.
- $\boldsymbol{w}$ le vecteur des charges de travail où $\boldsymbol{w_i}$ donne la charge de travail associée à la brique i.

1.2 Méthode et solveur

Voici le problème d'optimisation déduit de l'énoncé du projet :

$$
\begin{aligned}
\min_{x_{i,j} \in A} \quad & \sum_{j=1}^{4} \sum_{i=1}^{22} D_{i,j} \cdot x_{i,j} \\
\text{s.t.} \quad & \sum_{i=1}^{22} x_{i,j} \cdot \boldsymbol{w}_i \leq b_1, \quad j = 1..4 \\
& \sum_{i=1}^{22} x_{i,j} \cdot \boldsymbol{w}_i \geq b_2, \quad j = 1..4 \\
& \sum_{j=1}^{4} x_{i,j} = 1, \quad i = 1..22 \\
& \sum_{(i,j) \in BH} x_{i,j} = 4, \quad BH = \{(4,1), (14,2), (16,3), (22,4)\} \\
& b_1 \geq b_2 \geq 0
\end{aligned}
\tag{1}
$$

L'ensemble BH défini dans ce problème est l'ensemble des couples brique centrale/commercial, la contrainte mise sur cet ensemble peut se traduire comme : "chaque commercial doit être associé à sa brique centrale". Les constantes b_1 et b_2 sont les bornes utilisées pour encadrer la charge de travail, dans la première question du projet nous initialisons $b_1 = 1.2$ et $b_2 = 0.8$.

Ce problème ne minimise qu'un seul des deux objectifs, il faut aussi minimiser le nombre de réassignations. Ce nombre peut-être calculé en utilisant la matrice d'assignation initiale qui correspond à l'assignation actuel, nous notons cette matrice binaire A'. Ainsi, la fonction $f(x) = \frac{1}{2} \sum_{i=1}^{22} \sum_{j=1}^{4} (A_{i,j} - A'_{i,j})^2$ donne le nombre de réassignations pour une solution. Comme nous travaillons avec des matrices binaires, nous pouvons dire que :

$$
\begin{aligned}
(A_{i,j} - A'_{i,j})^2 \quad & = A_{i,j}^2 - 2A_{i,j}A'_{i,j} + {A'}_{i,j}^2 \\
& = A_{i,j} - 2A_{i,j}A'_{i,j} + A'_{i,j}
\end{aligned}
\tag{2}
$$

Nous pouvons désormais définir le problème d'optimisation associé au second objectif :

$$
\begin{aligned}
\min_{x_{i,j}\in A} \quad & \sum_{j=1}^{4}\sum_{i=1}^{22} x_{i,j} - 2x_{i,j}A'_{i,j} + A'_{i,j} \\
\text{s.t.} \quad & \sum_{i=1}^{22} x_{i,j}\cdot \boldsymbol{w}_i \leq b_1, \quad j = 1..4 \\
& \sum_{i=1}^{22} x_{i,j}\cdot \boldsymbol{w}_i \geq b_2, \quad j = 1..4 \\
& \sum_{j=1}^{4} x_{i,j} = 1, \quad i = 1..22 \\
& \sum_{(i,j)\in BH} x_{i,j} = 4, \quad BH = \{(4,1),(14,2),(16,3),(22,4)\} \\
& b_1 \geq b_2 \geq 0
\end{aligned}
\tag{3}
$$

Nous pouvons désormais trouver les solutions efficaces de ce problème multi-objectif en appliquant l'algorithme suivant :

- Trouver la solution optimale pour le premier problème en minimisant la distance.
- Compter le nombre de réassignations effectuées dans cette solution.
- Utiliser ce nombre comme une contrainte de borne supérieure dans le deuxième problème.
- Trouver la solution optimale pour le second problème en minimisant les réassignations.
- Calculer la distance totale pour cette solution.
- Utiliser la distance comme une contrainte de borne supérieure dans le premier problème.
- Répéter l'algorithme jusqu'à ce qu'il n'y ait plus de solution optimale trouvée.

Le solveur que nous utilisons pour calculer les solutions efficaces de ce problème provient de la bibliothèque Python OR-Tools, plus précisemment nous utilisons le solveur SCIP du module *pywraplp*. Nous aurions pu utiliser un sovleur du style CP-SAT, sachant que beaucoup de nos contraintes ne sont pas bornées par des nombres entiers cela nous aurait obligé à appliquer un facteur sur ces dernières. Vous trouverez le code que nous avons utilisé pour calculer ces solutions dans le fichier Jupyter Notebook "MIP_notebook.ipynb" (pensez à installer les modules nécessaires listés dans le fichier "requirements.txt").

2 Solutions au problème

Avant de présenter les différentes solutions que nous avons trouvées, voici la situation de base que nous allons tenter d'améliorer :

Commerciaux	Briques
1	[4,5,6,7,8,15]
2	[10,11,12,13,14]
3	[9,16,17,18]
4	[1,2,3,19,20,21,22]

Par la suite, nous appellerons "taux de partage des tâches" le nombre obtenu en faisant la différence entre la charge de travail maximale et la charge de travail minimale pour une solution. Nous pouvons arriver à un résultat similaire en faisant une moyenne des charges de travail sur une solution, cependant nous pensons que ce nombre est plus pertinent pour parler de solutions "équitables" puisqu'il montre clairement la différence entre le commercial travaillant le plus et le commercial travaillant le moins. En effet trouver des solutions équitables semble être une des approches du décideur car il est précisé que le modèle doit avoir des contraintes sur les charges de travail minimales et maximales. Plus le taux de partage des tâches est proche de 0, plus les tâches sont équitablement réparties entre les commerciaux. Plus ce nombre est élevé, plus il y a de disparités entre le commercial travaillant le plus et le commercial travaillant le moins. Voici les données de la solution courante concernant la distance totale parcourue et le taux de partage des tâches :

- La distance totale parcourue par les commerciaux est de 187.34km.
- Le taux de partage des tâches est de 0.6329.

2.1 Charge de travail entre 0.8 et 1.2

Nous avons pu trouver 7 solutions efficaces respectant les contraintes de charge de travail entre 0.8 et 1.2. Le minimum en distance est atteint par la solution 1 avec 154.62km au total, c'est aussi la solution avec le plus de réassignations. Le minimum en réassignations est atteint par la solution 2 avec 1 réassignation, cependant la solution possède la distance totale la plus élevée (182.2km). Voici le graphique donnant les solutions par rapport aux nombres de réassignations et à la distance totale :

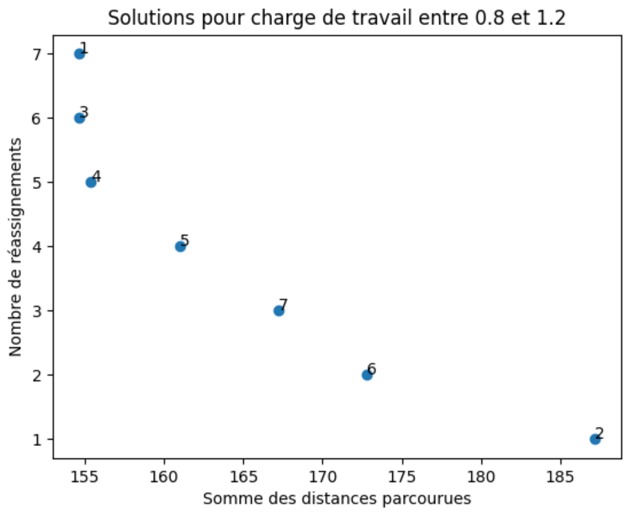

Figure 1: Ensemble des solutions entre 0.8 et 1.2 de charge de travail

Voici maintenant les données précises sur le total des distances, le nombre de réassignations et le taux de partage des tâches pour chaque solution.

Solution	Distance totale	Nbr réas.	Taux de partage des tâches
1	154.62	7	0.3121
2	187.2	1	0.1341
3	154.64	6	0.3247
4	155.41	5	0.3548
5	161.0	4	0.3206
6	172.8	2	0.2835
7	167.21	3	0.3519

Table 1: Tableau des distances totales et réassignations par solution

Vous trouverez ci-dessous les assignations des briques pour chaque solution.

Commerciaux	Briques
1	[4, 5, 6, 7, 8, 9, 12, 19, 20]
2	[11, 13, 14, 18]
3	[10, 15, 16, 17]
4	[1, 2, 3, 21, 22]

(a) Solution 1

Commerciaux	Briques
1	[4, 5, 6, 7, 8, 15]
2	[11, 12, 13, 14]
3	[9, 10, 16, 17, 18]
4	[1, 2, 3, 19, 20, 21, 22]

(b) Solution 2

Commerciaux	Briques
1	[4, 5, 6, 7, 8, 9, 19, 20]
2	[11, 12, 13, 14, 18]
3	[10, 15, 16, 17]
4	[1, 2, 3, 21, 22]

(c) Solution 3

Commerciaux	Briques
1	[4, 5, 6, 7, 8, 9, 19, 20]
2	[11, 12, 13, 14]
3	[10, 15, 16, 17, 18]
4	[1, 2, 3, 21, 22]

(d) Solution 4

Commerciaux	Briques
1	[4, 5, 6, 7, 8, 9, 21]
2	[11, 12, 13, 14]
3	[10, 15, 16, 17, 18]
4	[1, 2, 3, 19, 20, 22]

(e) Solution 5

Commerciaux	Briques
1	[4, 5, 6, 7, 8, 15, 21]
2	[11, 12, 13, 14]
3	[9, 10, 16, 17, 18]
4	[1, 2, 3, 19, 20, 22]

(f) Solution 6

Commerciaux	Briques
1	[4, 5, 6, 7, 8, 15, 19, 20]
2	[11, 12, 13, 14]
3	[9, 10, 16, 17, 18]
4	[1, 2, 3, 21, 22]

(g) Solution 7

Figure 2: Toutes les solutions trouvées par l'algorithme

2.1.1 Comparaison avec la solution courante

Pour rappel la solution courante possède une distance totale de 187.34km. Cela signifie qu'en ne faisant qu'une seule réassignation, la distance totale passe à 187.2km donc un gain de 0.14km. Ce gain peut être considéré comme négligeable, surtout qu'en augmentant le nombre de réassignations à 2 il y a un gain très clair sur la distance totale (14.54km). Bien sûr ce n'est pas notre rôle que de choisir à la place du décideur, il s'agit cependant d'une propriété intéressante à montrer lors de la présentation des solutions.

De plus nous pouvons remarquer que la distance totale change beaucoup entre deux solutions qui possèdent peu de réassignations, par exemple les solutions 2 et 6 font respectivement 1 et 2 réassignations cependant il y a un écart fort sur le total des distances (14.4km). A l'inverse quand le nombre de réassignations est grand les distances totales des solutions restent très proches. Les solutions 4 et 3 ne diffèrent que de 0.77km et la différence entre les solutions 1 et 3 est de 0.02km. C'est une chose à préciser au décideur qui pourrait être tenté de prendre la solution optimisant au mieux le total des distances ou le nombre de réassignations.

Toutes les solutions améliorent le taux de partage des tâches, pour rappel ce taux est à 0.6329 pour la solution courante. Le meilleur taux, et de loin, est celui de la solution 2 avec 0.1341. Par conséquent c'est la solution qui partage le mieux la charge de travail entre les commerciaux. Le deuxième meilleur taux est obtenu avec la solution 6 où il atteint 0.2835, ensuite tous les autres taux sont au dessus de la barre des 0.3. Si le partage équitable des tâches est une priorité du décideur, alors la solution 2 pourrait être une très bonne option malgré la distance totale qui reste très élevée comparée aux autres.

2.2 Charge de travail entre 0.9 et 1.1

Nous allons désormais nous intéresser aux solutions dont la charge de travail globale est comprise entre 0.9 et 1.1. Contrairement aux solutions précédentes le maximum de réassignations est à 6, la distance maximale reste la même avec 187.2km. Ces solutions sont au nombre de cinq, vous trouverez ci-dessous une représentation graphique des solutions :

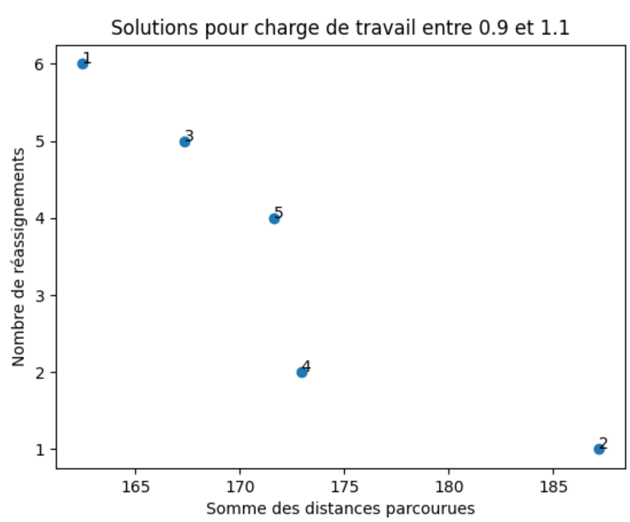

Figure 3: Ensemble des solutions entre 0.9 et 1.1 de charge de travail

Comme pour les solutions précédentes, voici le tableau regroupant pour chaque solution la distance totale, le nombre de réassignations et le taux de partage des tâches.

Solution	Distance totale	Nbr réas.	Taux de partage des tâches
1	162.43	6	0.1841
2	187.2	1	0.1341
3	167.33	5	0.1841
4	172.93	2	0.1823
5	171.62	4	0.1525

Table 2: Tableau des distances totales et réassignations par solution

De plus voici les assignations complètes pour chaque solution :

Commerciaux	Briques
1	[4, 5, 6, 7, 8, 9, 12, 19]
2	[11, 13, 14, 17]
3	[10, 15, 16, 18]
4	[1, 2, 3, 20, 21, 22]

(a) Solution 1

Commerciaux	Briques
1	[4, 5, 6, 7, 8, 15]
2	[11, 12, 13, 14]
3	[9, 10, 16, 17, 18]
4	[1, 2, 3, 19, 20, 21, 22]

(b) Solution 2

Commerciaux	Briques
1	[4, 5, 6, 7, 8, 9, 17, 19]
2	[11, 12, 13, 14]
3	[10, 15, 16, 18]
4	[1, 2, 3, 20, 21, 22]

(c) Solution 3

Commerciaux	Briques
1	[4, 5, 6, 7, 8, 15, 19]
2	[11, 12, 13, 14]
3	[9, 10, 16, 17, 18]
4	[1, 2, 3, 20, 21, 22]

(d) Solution 4

Commerciaux	Briques
1	[4, 5, 6, 7, 8, 15, 19]
2	[12, 13, 14, 18]
3	[9, 10, 11, 16, 17]
4	[1, 2, 3, 20, 21, 22]

(e) Solution 5

Figure 4: Toutes les solutions trouvées par l'algorithme

Nous pouvons de suite remarquer qu'une solution est commune aux deux intervalles : la solution 2 est en effet la même, les quatre autres solutions sont cependant différentes. Par ailleurs le modèle contraint à l'intervalle $[0.9, 1.1]$ ne possède pas de solution pour trois réassignations.

2.2.1 Comparaison avec la solution actuelle

La réelle amélioration apportée par toutes les solutions proposées est la diminution drastique du taux de partage des tâches. En effet, le taux maximum est de 0.1841 atteint par les solutions 1 et 3. La taux de partage des tâches minimum est quant à lui encore atteint par la solution 2 avec un taux à 0.1341. Le taux de partage des tâches reste très proche entre ces solutions avec un écart maximum de 0.05, le décideur pourrait très bien ignorer ce taux car il n'est pas assez discriminant entre ces solutions pour être utile.

Ces solutions sont cependant beaucoup moins efficaces par rapport à la distance totale, le minimum est atteint par la solution 1 avec 162.43km, le maximum reste la solution 2 avec 187.2km. Nous voulons préciser au décideur qu'il serait intéressant d'évaluer la solution 4 qui pour deux réassignations améliore bien mieux la distance totale parcourue que la solution 2 avec une seule réassignation.

2.3 Problématique de l'assignation partielle

Nous souhaitons vous prévenir de suite que nous ne sommes pas parvenu à implémenter cette partie. Nous avons cependant essayé et nous avons trouvé des idées que nous allons vous présenter ci-après pour résoudre ce problème.

L'assignation partielle veut dire qu'une brique n'est plus destinée qu'à un seul et même commercial, plusieurs commerciaux peuvent désormais s'en charger. Pour représenter cette situation, nous pouvons modifier les variables de décision pour en faire des réels compris entre 0 et 1. Quand la variable vaut 0, le commercial n'est pas assigné à cette brique, si la variable est inférieure à 1 alors le commercial partage cette brique avec d'autres commerciaux, et si la variable est égale à 1 alors le commercial est le seul responsable de cette brique. Prenons l'exemple d'assignation partielle suivant :

Brique	SR1	SR2	SR3	SR4
1	1	0	0	0
2	0.25	0.25	0.25	0.25
3	0.2	0.5	0	0.3

Table 3: Exemple d'assignation partielle

Dans cet exemple, la brique 1 est assignée exclusivement au commercial 1, la brique 2 est assignée équitablement entre les commerciaux, et enfin la brique 3 est assignée à 20% au commercial 1, 50% au commercial 2 et 30% au commercial 4 (le commercial 3 n'est pas assigné à cette brique).

Cette modélisation permet d'exprimer l'assignation partielle des briques, cela crée cependant plusieurs problèmes que nous allons préciser maintenant. Le premier point est que malgré une assignation de 0.1, le commercial assigné à une brique doit faire le trajet complet pour aller jusqu'à cette brique, par conséquent nous ne pouvons pas juste faire la somme sur les distances multipliées par l'assignation. Pour palier ce problème, il faudrait une fonction $f(x)$ telle que :

$$f(x) = \begin{cases} 0 & \text{si } x = 0 \\ 1 & \text{si } x > 0 \end{cases} \tag{4}$$

Cette fonction permet de résoudre le problème des distances mais elle n'est pas linéaire, il n'est donc pas possible de l'implémenter dans le solveur.

Le deuxième problème intervient dans l'optimisation du deuxième critère, celui du calcul des réassignations. Nous ne pouvons plus utiliser la technique suivante

$$\begin{aligned} (A_{i,j} - A'_{i,j})^2 &= A_{i,j}^2 - 2A_{i,j}A'_{i,j} + A'^2_{i,j} \\ &= A_{i,j} - 2A_{i,j}A'_{i,j} + A'_{i,j} \end{aligned} \tag{5}$$

En effet, ce ne sont plus des variables binaires par conséquent $(A_{i,j} - A'_{i,j})^2$ ne peut plus être simplifié de la sorte. Or cette fonction n'est pas linéaire et ne peut donc pas être utilisée dans un solveur. Nous pouvons cependant changer cette fonction par la fonction valeur absolue : $|A_{i,j} - A'_{i,j}|$. Cette fonction n'est toujours pas linéaires mais nous savons qu'il est possible de traduire cette fonction dans un contexte de programmation mathématique pour la remplacer par un équivalent linéaire.

2.4 Ajout d'un commercial

Nous nous intéressons désormais au cas où un nouveau commercial serait engagé car la charge de travail a augmenté de 20% pour chaque brique, la question de sa brique centrale est mise en avant.

Nous pouvons garder les mêmes définitions utilisées précédemment pour les deux problèmes d'optimisation, la différence se fera sur l'indice j qui prendra ses valeurs entre 1 et 5. Cependant, il nous manque une information cruciale pour la résolution de ce problème : la matrice complète des distances entre toutes les briques. En effet nous utilisons pour l'instant une matrice des distances réduite qui ne décrit que les distances entre les briques centrales et les autres briques, sachant que le cinquième commercial n'a pas de brique centrale nous ne pouvons pas utiliser cette matrice pour ce problème. Dès que la matrice des distances entre toutes les briques est fournie, nous pouvons relancer la résolution du problème d'optimisation pour connaître les meilleures assignations pour le nouveau commercial.

Pour chaque solution, nous devons désigner la brique centrale du nouveau commercial. Pour cela, nous devons prendre les briques assignées au nouveau commercial et pour chaque brique sommer les distances de cette brique à toutes les autres. Ainsi nous pouvons prendre la brique qui possède la distance totale la plus faible et l'utiliser comme brique centrale.

2.5 Modification des briques centrales

Nous nous intéressons maintenant au cas où les commerciaux peuvent modifier leurs briques centrales. Nous avons pensé à un moyen simple et intuitif pour créer à la fois les assignations et pour trouver les briques centrales adéquates : utiliser la méthode k-means.

L'algorithme k-means est fait pour créer un partitionnement d'un ensemble de points (ou vecteurs) en k clusters de telle sorte à ce que les points proches se trouvent dans le même cluster. C'est exactement ce que l'on souhaite ici : nous souhaitons partitionner un ensemble de points (briques) en 4 clusters (assignations). Pour que l'algorithme fonctionne, il nous faut la position des briques dans l'espace : nous pouvons approximer cette information en se basant sur la carte pour donner des coordonnées ad-hoc aux briques. Ce n'est pas l'idéal, nous aurions préféré par exemple avoir les coordonnées GPS précises de chaque brique mais il nous suffit que les coordonnées respectent la cohérence du plan pour que l'algorithme fonctionne et donne des résultats cohérents. L'algorithme se déroule de la manière suivante :

- Créer aléatoirement une partition des briques en k ensembles
- Calculer le barycentre de chaque cluster, c'est-à-dire calculer les coordonnées du point moyen au cluster (ce point n'est pas forcément une brique, c'est un point artificiel qui servira pour la suite)
- Pour chaque brique calculer $d_i \cdot w_i$ soit la distance de la brique au barycentre pondérée par la charge de travail de la brique et choisir le barycentre qui minimise ce nombre. Ajouter la brique au cluster associé au barycentre.
- Refaire le calcul des barycentres avec les nouveaux clusters et répéter le processus de réassignation tant que les clusters sont modifiés entre deux itérations.
- Après la dernière itération, pour chaque cluster choisir la brique qui minimise la distance totale à toutes les autres briques du cluster et désigner cette brique comme brique centrale pour le commercial associé au cluster.

Cet algorithme a pour grand avantage de toujours converger vers une solution. Nous voyons cependant deux grands défauts à utiliser cet algorithme ici : cela nécessite une création de données pour le faire fonctionner (les coordonnées sur le plan), et de plus nous n'avons pas de garantie que la solution soit optimisée entre certains seuils de charge de travail.

2.6 Préférences entre les briques

Finalement, nous allons étudier la question de l'implémentation des préférences des commerciaux entre les briques. Nous allons supposer que les commerciaux possèdent chacun un préordre total sur les briques pour pouvoir les comparer. A partir de chaque préordre, nous pouvons déduire une valuation des briques telle que : plus la brique possède une valuation proche de 0, moins la brique est préférée par le commercial. Nous pouvons implémenter cela sous la forme d'une matrice dont les lignes représentent les préférences valuées des commerciaux pour toutes les briques.

Ensuite nous devons savoir ce que nous souhaitons calculer : nous pouvons associer une assignation à un "niveau de bonheur" des commerciaux, c'est-à-dire à quel point l'assignation a pris en compte les préférences de chaque commerciaux. Il existe plusieurs méthodes pour cela (méthode égalitariste, utilitariste, méthode de Nash...), ici nous allons utiliser la méthode utilitariste c'est-à-dire que l'on souhaite maximiser la somme des valuations de tous les commerciaux. Soit A une matrice d'assignation, soit L la matrice des préférences valuées où $L_{j,i}$ désigne la valuation de la brique i pour le commercial j. La méthode utilitariste définit qu'un bon assignement doit maximiser la fonction suivante :

$$\sum_{j=1}^{4} \sum_{i=1}^{22} A_{i,j} \cdot L_{j,i}$$

. Nous pouvons voir cela de deux façons :

- Comme un objectif : nous nous retrouvons avec un problème tri-objectif distance, réassignations, préférences. La distance et les réassignations sont a minimiser, et la fonction utilitariste est à maximiser (on souhaite maximiser le bonheur des commerciaux).
- Comme une contrainte : le décideur décide d'un seuil arbitraire qui servira de borne inférieure à la fonction utilitariste. Nous restons donc sur un problème bi-objectif.