PIAULAGE DES ADMISSIBLES DANS LA RÉSIDENCE MEUNIER

LE CHENADEC BASTIEN & TERRISSE THEÏLO

Encadrant : VINCENT LECLÈRE

École des Ponts ParisTech

CLUB INFORMATIQUE

2021-2022

PRÉSENTATION DU PROBLÈME

Cette année encore, les admissibles au concours Mines-Ponts passant leurs oraux aux Ponts pourront demander à être logés dans la résidence Meunier. Pour cela, le *KI*, club informatique des Ponts, a développé le site « **Admissibles** » [1] (figure 1), permettant aux étudiants de formuler leurs demandes. L'objet de ce projet est l'étude de ce problème d'attribution afin que le site répartisse automatiquement les élèves dans les chambres.

La résidence propose un nombre limité de **chambres de 3 types : simple, binomée ou double**. Une chambre simple ne peut accueillir qu'une personne, les autres 2 personnes. Chaque semaine est organisé un **« shotgun »**, à l'occasion duquel les élèves formulent leur demande, un élève s'inscrivant forcément pour une semaine complète. On modélise le problème à partir des contraintes suivantes :

- **critères primaires :** il s'agit par ordre d'importance de la **distance**[2] entre le domicile et l'école, du **caractère boursier** ou non puis du **rang dans le shotgun**;
- **critères secondaires :** un candidat peut formuler une **préférence en termes de type de chambre**, et deux élèves peuvent explicitement **demander à être en colocation**;

FIGURE 1 – Page d'accueil du site Admissibles

- **contraintes : de genre**, deux candidats de genres opposés[3] ne pouvant pas être colocataires à moins de l'avoir demandé; **de préférence «stricte»**, un candidat pouvant refuser une chambre si elle n'est pas du type qu'il préfère.

Ces critères, et leur contribution dans les attributions, sont quantifiés par des cœfficients (B_i pour un bonus, P_i pour un malus) dont les valeurs absolues sont recensées dans le tableau 1.

$\mathbf{B_f}$ ($> 800km$)	$\mathbf{B_d}$ ($> 50km$)	$\mathbf{B_s}$ (bourse)	$\mathbf{P_r}$ (shotgun)	$\mathbf{B_p}$ (pref. chambre)	$\mathbf{B_b}$ (colocation)
1	0.3	0.2	1×10^{-4}	1×10^{-6}	1×10^{-3}

TABLE 1 – Valeurs des cœfficients

[1] https://admissibles.enpc.org

[2] On distingue pour le critère de distance 3 paliers séparés par 2 seuils : $50km$ et $800km$.

[3] À noter qu'aucune contrainte n'est appliquée si l'étudiant ne précise pas son genre.

CONSIDÉRATIONS THÉORIQUES

Cas simples

On commence par étudier 4 cas simples. On note n le nombre de demandes et m_S et m_D les nombres de chambres simples et doubles respectivement.

Uniquement des chambres simples ($m_S \in \mathbb{N},\ m_D = 0$) : On affecte les demandes par ordre décroissant de scores primaires.
→ Complexité en $\boxed{O(n\log(n))}$.

Uniquement des chambres doubles ($m_S = 0,\ m_D \in \mathbb{N}$) : On teste toutes les solutions (tous les partitionnements de $2m_D$ demandes en m_D paires).
→ Complexité en $\boxed{O(\binom{n}{2m_D})(2m_D-1)(2m_D-3)...1)}$.

Uniquement des simples et une double ($m_S \in \mathbb{N},\ m_D = 1$) : On teste, pour chaque paire d'élèves placée dans la chambre double, la résolution des autres demandes avec seulement des chambres simples.
→ Complexité en $\boxed{O(\binom{n}{2}) \times C_{simple}(m_S, n-2)) = 0(n^3\log(n))}$

(avec C_{simple} la complexité pour des chambres simples uniquement).

Uniquement des doubles et une simple ($m_S = 1,\ m_D \in \mathbb{N}$) : On teste, pour chaque élève placé dans la chambre simple, la résolution des autres demandes avec seulement des chambres doubles.
→ Complexité en $\boxed{O(n \times C_{double}(m_D, n-1))}$

Ces cas simples montrent que la difficulté du problème réside dans les interactions entre colocataires, et qu'une approche exacte semble peu fructueuse par des méthodes naïves. Ainsi, **nous décidons d'utiliser un PLNE**.

Complexité du problème

L'apparente difficulté du problème, même pour des cas simples, invite à en étudier la complexité. Or, on montre que le problème peut être mis sous la forme:

$$\begin{aligned} \max_{x} \quad & \frac{1}{2}x^T\mathbf{Q}x + c^Tx \\ \text{s.c.} \quad & \mathbf{A}x \leqslant b \end{aligned}$$

où $\mathbf{Q}$ est une matrice symétrique réelle qui n'est ni définie négative, ni semi-définie négative : le relâché continu de notre problème appartient donc à la classe des problèmes d'optimisation quadratique non convexe, réputés difficiles.

Borne supérieure sur la valeur optimale

Pour obtenir une borne supérieure pour notre problème de maximisation et se faire une idée de la qualité de la solution, nous relâchons les contraintes de genre et de préférence stricte. Pour une instance dont les chambres offrent une capacité totale C_t, on trie alors les demandes par ordre de scores primaires décroissants et on ne retient que les C_t meilleures. Enfin, la taille réduite de cette nouvelle instance permet une résolution en temps raisonnable.
Le score retourné est bien une borne supérieure de la solution optimale dès lors que la perte générée par le remplacement d'une demande sélectionnée par une autre de moindre score primaire n'est pas compensée par le gain engendré par l'amélioration des critères secondaires, ce que permet le choix des valeurs des paramètres (*cf* tableau 1).

PLNE

On introduit les variables binaires x_{ij} et $z_{i_1i_2j}$:

$$x_{ij} = \begin{cases} 1 & \text{si la demande } i \text{ reçoit} \\ & \text{ la chambre } j \\ 0 & \text{sinon} \end{cases} \qquad z_{i_1i_2j} = \begin{cases} 1 & \text{si les demandes } i_1 \text{ et } i_2 \\ & \text{ reçoivent la chambre } j \\ 0 & \text{sinon} \end{cases}$$

Le problème linéaire en nombres entiers s'écrit :

$$\max_{x,z} \quad \sum_{i,j} x_{ij}\left(1 + \overbrace{B_d\mathbb{1}^{(i)}_{>50km}}^{\text{bonus distance}} + \overbrace{B_f\mathbb{1}^{(i)}_{>800km}}^{\text{bonus grande distance}} + \overbrace{B_s\mathbb{1}^{(i)}_{\text{bourse}}}^{\text{bonus boursier}} - \overbrace{P_rr_i}^{\text{malus shotgun}}\right)$$
$$+ \sum_{i,j} x_{ij} \overbrace{B_pp_{ic_j}}^{\text{bonus préférence de chambre}} + \sum_{i_1<i_2,j} z_{i_1i_2j} \overbrace{B_bb_{i_1i_2}}^{\text{bonus colocation}}$$

$$\text{s.c.} \begin{cases} \forall(i,j), & x_{ij} \in \{0,1\} \\ \forall i, & \sum_j x_{ij} \leqslant 1 & \textit{(Au plus une chambre par élève)} \\ \forall j, & \sum_i x_{ij} \leqslant \texttt{capacité}_j & \textit{(Capacité des chambres)} \\ \forall i_1 < i_2, \forall j, & \begin{cases} z_{i_1i_2j} \geqslant 0 \\ z_{i_1i_2j} \leqslant x_{i_1j} \\ z_{i_1i_2j} \leqslant x_{i_2j} \\ z_{i_1i_2j} \geqslant x_{i_1j} + x_{i_2j} - 1 \end{cases} & \textit{(Linéarisation de la contrainte } z_{i_1i_2j} = x_{i_1j}x_{i_2j}\textit{)} \\ \forall j \text{ tq } \texttt{capa}_j > 1, & \sum_{i_1<i_2} z_{i_1i_2j}|g_{i_1}g_{i_2}(g_{i_1} - g_{i_2})|(1 - b_{i_1i_2}) = 0 & \textit{(Genres)} \\ \forall(i,j), & (1 - p_{ic_j})q_ix_{ij} = 0 & \textit{(Préférence stricte)} \end{cases}$$

HEURISTIQUE

Le PLNE donne une solution exacte mais ne fonctionne en temps raisonnable que sur des petites instances. On présente ici une heuristique pour des instances non résolubles par le PLNE, qui consiste à construire des groupes de demandes et de chambres qu'on résout séparément.

Algorithme 1 Heuristique

```
Entrée: demandes, chambres
  Trier les demandes selon le score primaire.
  tant que il reste des chambres faire
     Former un groupe avec les 50 premières demandes non attribuées.
     Tant qu'il reste des chambres, en attribuer suffisamment au groupe
     pour que chaque élève du groupe ait une place.
  fin tant que
  pour chaque groupe faire
     Résoudre avec le PLNE le sous-problème composé des demandes et
     des chambres attribuées au groupe.
  fin pour
  Construire une solution en agrégeant toutes les solutions aux sous-problèmes.
  Appliquer un algorithme de recherche locale pour améliorer la solution. On
  utilise comme voisinage des échanges d'étudiants (4 types d'échanges
  implémentés).
```

En théorie, on résout $\lceil n/50 \rceil$ problèmes de complexité bornée avec le PLNE, où n est le nombre de demandes : complexité linéaire avec une très grande constante multiplicative. De plus, chaque opération de la recherche locale a une complexité en pire quadratique, donc la complexité totale de l'heuristique est quadratique.

DISCUSSION

- La formulation du PLNE peut être améliorée en brisant certaines symétries : des solutions de même score peuvent coexister, par exemple par permutation de deux élèves en chambre simple.
- Le problème gagnerait peut-être à être formulé sous la forme d'un problème d'optimisation multi-objectif.
- Certains choix arbitraires ont été réalisés dans la formulation du problème. Par exemple, une alternative aux préférences simples aurait pu être un classement des types de chambres préférés par l'étudiant.
- Le problème peut être complexifié si sont intégrées des contraintes de cohérence des arrivées et départs des étudiants : le temps s'ajoute à la formulation du PLNE.
- L'heuristique peut encore être améliorée en augmentant la taille des groupes; en testant des groupes différents et en ne gardant que les meilleurs (les chambres sont attribuées aléatoirement aux groupes); ou en laissant tourner la recherche locale plus longtemps.

RÉSULTATS

Demandes	Chambres	Temps d'exécution PLNE (s)	Temps d'exécution heuristique (s)	Score PLNE	Score heuristique	Score algorithme de référence	Borne supérieure
50	11	1	2	36	36	34	36
100	31	18	9	90	90	80	90
200	51	194	20	152	149	144	154
300	71	477	48	212	212	199	216
400	91	Dépassement mémoire (64Go)	109	-	260	253	273
500	111	Dépassement mémoire (64Go)	141	-	342	256	342

TABLE 2 – Résultats obtenus sur différentes instances générées aléatoirement.

Comparaison avec un algorithme de référence

On propose de comparer notre heuristique avec l'algorithme utilisé sur le site de l'année 2020-2021 (référence) qui consiste à :

1. Trier les demandes par critère primaire.
2. Pour chaque demande, essayer de placer l'élève dans une chambre où il reste une place sans enfreindre les contraintes de genre ni celles de préférence stricte, sinon refuser la demande.

Sur l'instance de 500 demandes générées aléatoirement, on obtient un score de référence de 255 contre 341 pour l'heuristique.

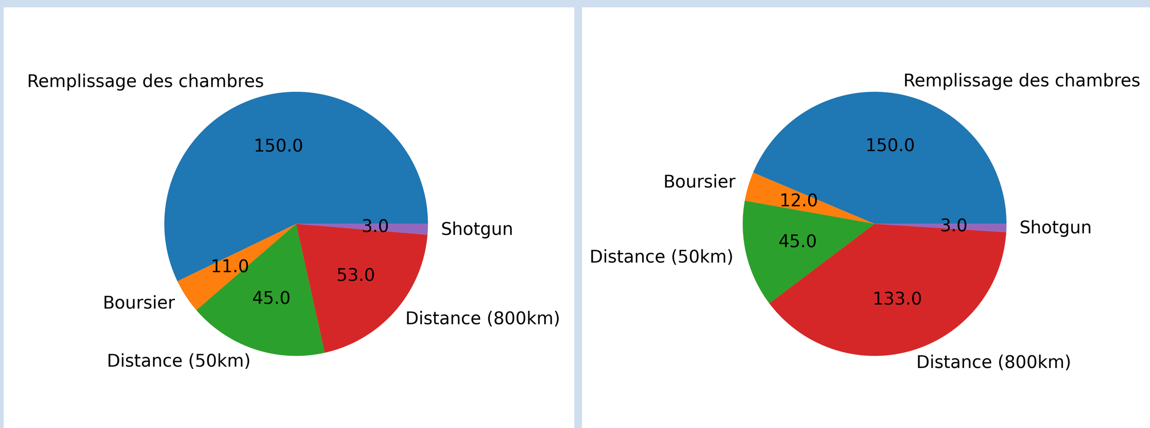

FIGURE 2 – Répartition du score entre les critères primaires (algorithme de référence à gauche, heuristique à droite)

L'algorithme de référence donne un bon score, mais l'heuristique est environ un tiers meilleure. La figure 2 montre que la meilleure répartition a permis de placer des élèves avec un score primaire élevé que l'algorithme de référence n'a pas réussi à placer, l'obligeant à choisir des élèves de moins bon score primaire.

Analyse de la solution

L'heuristique est testée sur l'instance de l'année 2020-2021, comportant 522 demandes ainsi que 72 chambres simples, 23 chambres binomées et 16 chambres doubles pour une capacité totale de 150 places.

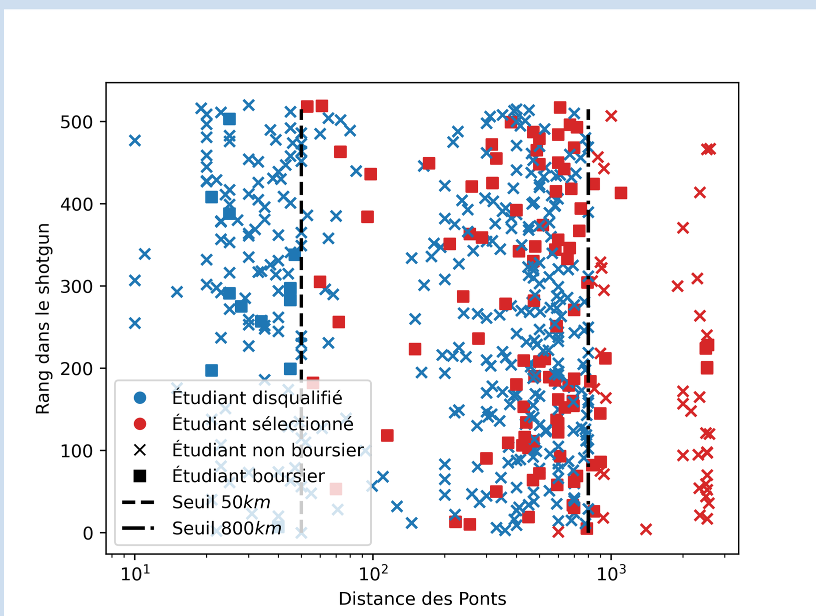

FIGURE 3 – Synthèse des critères primaires et de leur sélection

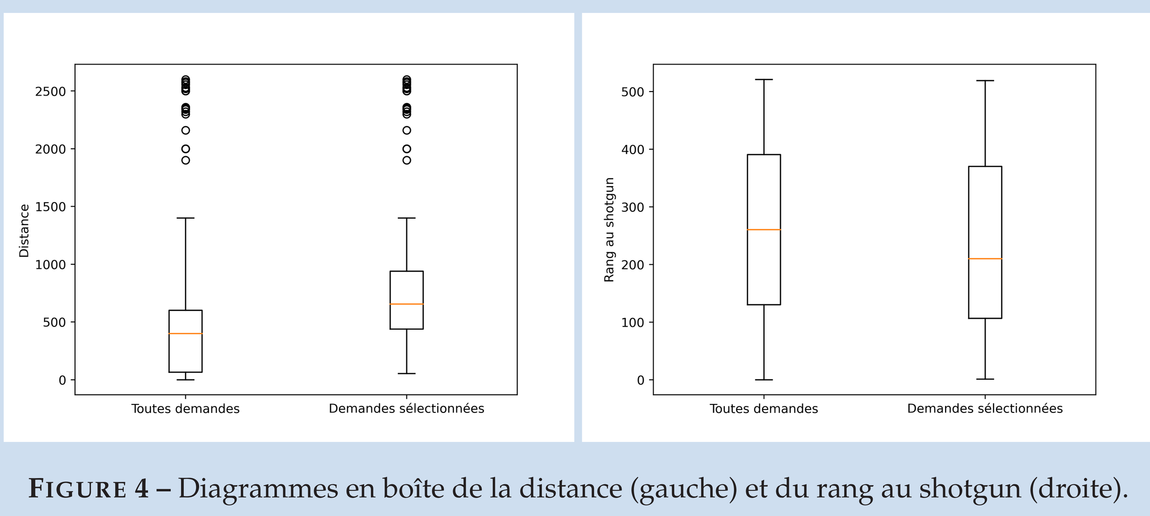

FIGURE 4 – Diagrammes en boîte de la distance (gauche) et du rang au shotgun (droite).

La figure 3 montre bien une attribution priorisant les élèves très distants, puis distants, boursiers et, dans une moindre mesure, de faible rang au shotgun. Les diagrammes en figure 4 montrent aussi que les élèves distants sont privilégiés.
Notons enfin que les rangs au shotgun moyen et médian parmi les élèves sélectionnés sont de 238 et 210 respectivement, ce qui montre l'importance modérée de ce critère.

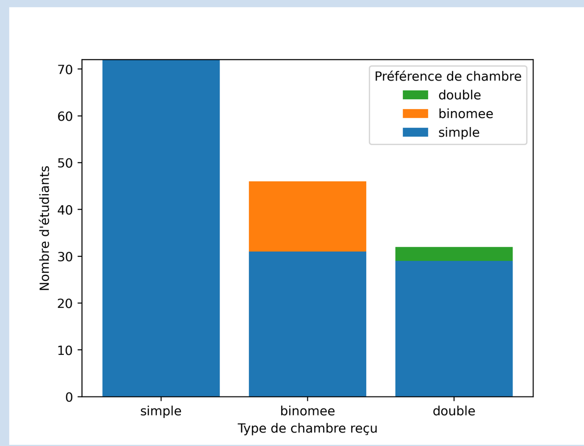

FIGURE 5 – Distribution des préférences dans chaque type de chambre

La figure 5 montre que l'algorithme satisfait au mieux les préférences de chambre, sachant que beaucoup de personnes ont émis une préférence pour la chambre simple. Enfin, l'heuristique respecte bien les demandes de colocation parmi les étudiants ayant eu une chambre.

Comparaison heuristique et PLNE

Pour comparer l'heuristique et la solution optimale, on utilise une instance résoluble par le PLNE (300 demandes). On note que :

- Les répartitions des types de demandes dans les différentes chambres sont très similaires.
- Les critères de bourse, de distance et de shotgun mènent à une sélection très similaire des élèves. Toutefois l'heuristique prend moins en compte le rang dans le shotgun.
- L'erreur est de l'ordre de 1%.

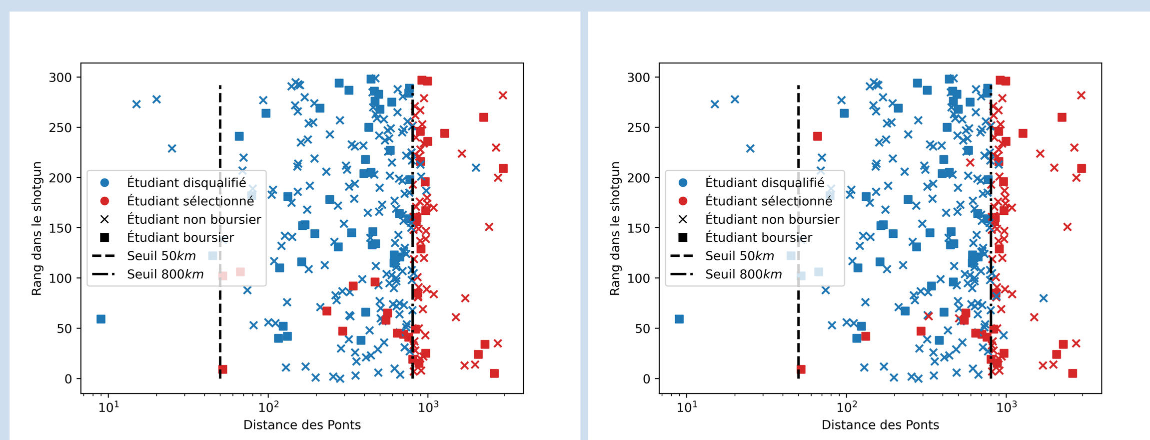

FIGURE 6 – Comparaison des critères primaires (PLNE à gauche, heuristique à droite)

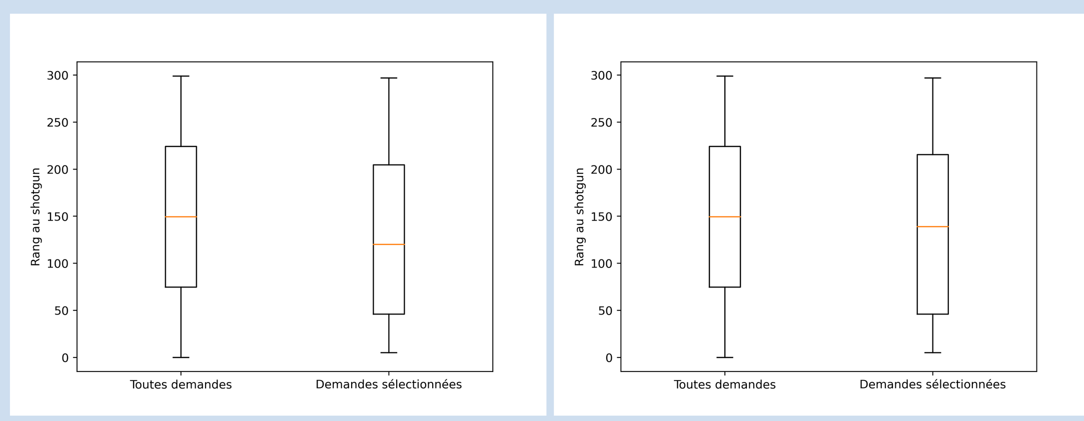

FIGURE 7 – Prise en compte du shotgun (PLNE à gauche, heuristique à droite)