

Examen “Conduite de projets en science des données et Predictive Analytics”

Question 1 : IA

Décrire la différence entre approche réaliste et approche utilitariste dans la démarche scientifique.

Il existe trois mondes : le monde des connaissances, le monde des idées et le monde des informations. Dans l'approche réaliste (ou platonicienne), on considère que le but de la science est de créer de la connaissance et d'explorer le monde des idées alors que dans l'approche utilitariste, on considère que la science doit juste savoir à expliquer des phénomènes, sans se soucier des concepts cachés derrière.

Question 2 : Programmation logique / chaînage avant

Dans un langage à base de règles simple en chaînage avant (on appelle cela un système de production) on a le programme:

```
var input[][], result[], i=1, tmp=0;
when input.length>0 and i >= input.length then result.append(input[tmp]),
input.removeAt(tmp), i=1, tmp=0;
when input[i] < input[tmp] then tmp=i, i=i+1;
when input[i] >= input[tmp] then i=i+1;
```

Lorsque l'instruction `input=[2,0,5,4,9];` est exécutée, que contiendront les variables `result` et `input` en retour? Expliquer l'algorithme. Donner sa complexité.

En fin d'exécution, les vecteurs sont les suivants : `result = [0;2;4;5;9]` et `input = []`.

L'algorithme parcourt la liste `input` pour trouver la valeur minimale (située au rang `tmp`).

Lorsque la liste a été parcourue entièrement, la valeur minimale est stockée dans la liste `result` et supprimée de la liste `input`.

De part le parcours de liste, la complexité est linéaire.

https://fr.wikipedia.org/wiki/Analyse_de_la_complexit%C3%A9_des_algorithmes

Question 3 : Smart City

Trouver sur internet 3 logiciels commerciaux professionnels destinés à remplir un rôle similaire à celui du projet SmartDeliveries. En vous basant sur la présentation commerciale, identifiez leurs principales caractéristiques démarquantes (quelles fonctionnalités mettent-ils particulièrement en avant par rapport à la concurrence). Fournir les références utilisées.

- Mapotempo : Géocodage d'adresses, calcul d'itinéraires, optimisation et suivi d'exécutions de tournées. <https://www.mapotempo.com/>

- ERP COPILOTE par Infologique : visualisation graphique du parcours, réduction des frais de transport, édition des feuilles de route, application sur tablette pour le chauffeur. <https://www.youtube.com/watch?v=A-gfJ5KPgSM>
- AntsRoute : Planification “intelligente” de tournées, facilité d'utilisation, suivi en temps réel, interactions avec le client, adaptation à tous types de livraison. <https://antsroute.com>

Question 4: Trafic routier

a- *Quelles sont les principales variables mesurées par un détecteur de trafic ?*

La date et l'heure, le débit par heure (nombre de véhicules qui passent en une heure), le taux d'occupation. On en déduit la densité du trafic et la congestion, qui peut être due à trois types de causes : événements connus ou anticipés, saturation chronique ou non du réseau et accidents.

b- *Qu'est-ce que le diagramme fondamental d'un détecteur de trafic, pourquoi est-il utile pour mesurer et prévoir la congestion ?*

Le diagramme fondamental d'un détecteur de trafic est un graphique qui représente la congestion (le taux d'occupation de la route) en fonction du temps. On peut définir grâce à lui le seuil à partir duquel on va observer que les voitures ralentissent lorsque le débit augmente, ainsi que le seuil à partir duquel le nombre de véhicules est si élevé que le débit reste constant (car ils sont arrêtés).

c- *Quel est le débit typique maximal d'un tronçon à une voie ?*

Le débit est plus important sur voie rapide ou autoroute qu'en zone urbaine, de part la présence de virages et d'issues plus nombreuses.

Question 5: Temps de parcours

a- *Quelles sont les principales variables prédictives du temps de parcours d'un camion de livraison en ville, par ordre d'importance décroissante ?*

Les principales variables sont : la distance à parcourir, l'identifiant de la tournée qui comprend l'expérience du conducteur, essentielle, et la prévision du temps de trajet statique.

b- *Citer 2 facteurs potentiels affectant les temps de parcours et difficiles à mesurer avec les données fournies dans les fichiers fournis en TP.*

Les adresses incomplètes et l'expérience du chauffeur qui peut savoir que le chargement doit se faire à l'arrière du magasin par exemple, donc pas exactement à l'adresse indiquée

Question 6. Prescriptive Analytics

Les affirmations suivantes sont-elles vraies ou fausses ?

a- *Un problème de décision est dans la classe de complexité NP si et seulement s'il n'existe pas d'algorithme polynomial pour le résoudre.*

Faux : Un problème de décision est dans la classe de complexité NP s'il est vérifié par un vecteur X ou s'il n'existe pas d'algorithme polynomial pour le résoudre. Il existe donc des

solutions vérifiées par des algorithmes polynomiaux dans la classe de complexité NP. C'est notamment le cas des problèmes appartenant à la classe P, inclue dans NP.

b- Dans l'industrie, la majorité des problèmes d'ordonnancement sont résolus grâce à des heuristiques.

Faux : ils sont résolus par des approches génériques et de la programmation par contraintes.

c- Le problème suivant possède exactement trois solutions:

$u \in \{1,3\}$; $v \in \{1,2\}$; $w \in \{3,4\}$; $x \in \{1,5\}$; $y \in \{4,5\}$ and $\text{allDifferent}(u,v,w,x,y)$

Faux : comme le montre le tableau suivant matérialisant un arbre de décision, seules deux solutions existent à ce problème : (1;2;3;5;4) et (3;2;4;1;5).

| u | v | w | x | y |
|---|---|---|---|---|
| 1 | 1 | 3 | 1 | 4 |
| | | | 1 | 5 |
| | | | 5 | 4 |
| | | | 5 | 5 |
| | 2 | 4 | 1 | 4 |
| | | | 1 | 5 |
| | | | 5 | 4 |
| | | | 5 | 5 |
| 3 | 1 | 3 | 1 | 4 |
| | | | 1 | 5 |
| | | | 5 | 4 |
| | | | 5 | 5 |
| | 2 | 4 | 1 | 4 |
| | | | 1 | 5 |
| | | | 5 | 4 |
| | | | 5 | 5 |

d- L'algorithme de résolution de CP-Optimizer est un algorithme exact: si un problème d'optimisation est faisable, il garantit de trouver une solution optimale.

Vrai : l'algorithme est exact.

Question 7. Optimisation

a- En cherchant sur internet, décrivez un problème d'optimisation combinatoire non vu dans le cours dont la version de décision est un problème NP-Complet.

Le problème du sac à dos qui consiste à remplir un sac avec des objets qui ont un poids et une valeur donnés. Le but est d'emporter la valeur maximale sans dépasser un poids fixé.

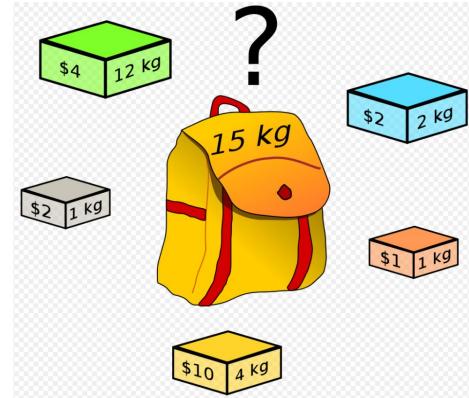
https://fr.wikipedia.org/wiki/Probl%C3%A8me_du_sac_%C3%A0_dos

https://fr.wikipedia.org/wiki/21_probl%C3%A8mes_NP-complets_de_Karp

b- Décrivez une petite instance particulière de ce problème d'optimisation (avec des valeurs pour chacune des données).

L'exemple donné par Wikipédia a été choisi : le sac à dos ne supporte pas plus de 15 kg et on a à notre disposition cinq poids tels que :

| Numéro du poids | Valeur (\$) | Poids (Kg) |
|-----------------|-------------|------------|
| 1 | 4 | 12 |
| 2 | 2 | 1 |
| 3 | 10 | 4 |
| 4 | 1 | 1 |
| 5 | 2 | 2 |



c- Donnez une solution faisable non-optimale et une solution optimale de cette petite instance.

En reprenant la même méthode que précédemment, on calcule toutes les solutions possibles à l'aide d'un tableur. Pour trouver une solution faisable non-optimale, on élimine les solutions dont le poids total excède 15 kg et on en choisit une dont la valeur n'est pas maximale. La solution optimale sera celle dont le poids sera inférieur à 15 kg et dont la valeur totale est maximale. On obtient qu'une solution possible est de prendre le poids de 12 kg et les deux de 1 kg, pour un poids total de 14 kg et une valeur de 7\$ (surlignée en jaune). La solution optimale est de prendre tous les poids sauf celui de 12 kg, le poids total est alors de 8kg, pour une valeur de 15\$ (surlignée en vert).

N.B : dans le tableur sont juste écrits les poids et pas les valeurs pour faciliter sa lecture.

| Poids 1 | Poids 2 | Poids 3 | Poids 4 | Poids 5 | Somme poids | Somme valeur |
|---------|---------|---------|---------|---------|-------------|--------------|
| 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | | 2 | 2 | 2 | 2 |
| | | | 1 | 0 | 1 | 1 |
| | | | 2 | 3 | 3 | 3 |
| | 4 | 0 | 0 | 4 | 4 | 10 |
| | | | 2 | 6 | 6 | 12 |
| | | | 0 | 5 | 5 | 11 |
| | | | 1 | 2 | 7 | 13 |
| | 1 | 0 | 0 | 1 | 1 | 2 |
| | | | 2 | 3 | 3 | 4 |
| | | 1 | 0 | 2 | 2 | 3 |
| | | | 2 | 4 | 4 | 5 |
| | | 4 | 0 | 5 | 5 | 12 |
| | | | 2 | 7 | 7 | 14 |
| | | | 0 | 6 | 6 | 13 |
| | | | 1 | 2 | 8 | 15 |
| 12 | 0 | 0 | 0 | 12 | 12 | 4 |
| | | | 2 | 14 | 14 | 6 |
| | | | 1 | 13 | 13 | 5 |
| | | | 2 | 15 | 15 | 7 |
| | 4 | 0 | 0 | 16 | 16 | 14 |
| | | | 2 | 18 | 18 | 15 |
| | | | 0 | 17 | 17 | 15 |
| | | | 1 | 19 | 19 | 17 |
| | 1 | 0 | 0 | 13 | 13 | 6 |
| | | | 2 | 15 | 15 | 8 |
| | | 1 | 0 | 14 | 14 | 7 |
| | | | 2 | 16 | 16 | 9 |
| | | 4 | 0 | 17 | 17 | 16 |
| | | | 2 | 19 | 19 | 18 |
| | | | 0 | 18 | 18 | 17 |
| | | | 1 | 20 | 20 | 19 |

Question 8. Programmation par contraintes

Deux principes fondamentaux de la Programmation par Contraintes sont (1) la recherche arborescente et (2) le filtrage du domaine des variables. Décrivez brièvement ces principes, leurs rôles et la façon dont ils sont mis en œuvre durant la résolution.

La recherche arborescente consiste à fixer les variables de décision une par une. On explore toutes les valeurs possibles de chaque branche. En cas de problème, on revient au noeud précédent pour tester d'autres solutions filles. Dans l'exemple du sudoku, cela revient à fixer la première case sur le 1, la suivante sur le 2, etc. Le but est donc de tester l'ensemble des possibilités (chaque nœud) pour arriver à la solution. Par conséquent, cette méthode n'est pas optimale.

De l'autre côté, le filtrage permet d'enlever des valeurs (donc des nœuds) car une contrainte existe et indique l'impossibilité de prendre cette valeur (pour un sudoku, une case ne prendra pas un chiffre s'il est déjà sur la ligne, la colonne ou la case correspondante par exemple). Le but est de simplifier le problème, non pas de le résoudre.

<http://ufrsciencestech.u-bourgogne.fr/m2bdia/UE7%20Outils%20de%20l'IA/PPC/Archives/CM/cours-ppc-M2.pdf>