

# Examen ' Conduite de projets en science des données et Predictive Analytics'

Craspay Lucie

## 1: IA

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

Une approche utilitariste est une forme de biomimétisme pour aborder des problèmes complexes. Il s'agit d'une métaphore pour construire des modèles de la pensée humaine, les interpréter et les généraliser. Approche utilisée en IA.

Une approche réaliste doit permettre d'expliquer comment les sciences parviennent avec succès à prédire les phénomènes, fondements épistémologiques, à apprêhender la complexité en cherchant à comprendre ce qui fonctionne, pour qui, et dans quels contextes. Une approche réaliste énonce qu' "une théorie scientifique acceptée et consolidée décrit exactement comment les choses se passent dans le monde".

Donc ok pour une approche utilitariste de la science, pas pour une approche réaliste.

## 2: programmation logique/chainage 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.

`result=[0,2,4,5,9] input=[ ]`

Il s'agit d'un algorithme pour ranger une suite par ordre croissant. On récupère la plus petite valeur, on la met dans tmp puis on supprime cette valeur de la liste input.

Tri par insertion. Complexité :  $O(n^2)$  en moyenne et dans le pire des cas,  $O(n)$  dans le meilleur des cas

Il s'agit du tri souvent utilisé naturellement pour trier des cartes à jouer : les valeurs sont insérées les unes

après les autres dans une liste triée (initialement vide). C'est souvent le plus rapide et le plus utilisé pour trier des entrées de petite taille. Il est également efficace pour des entrées déjà presque triées.

**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.

Le projet SmartDeliveries vise à optimiser la mobilité et combattre la saturation des routes en centre ville en prévoyant le trafic à l'avance, en prévoyant les déplacements des véhicules, en optimisant les chemins à emprunter.

D'autres logiciels commerciaux professionnels ont des rôles similaires à celui de SmartDeliveries:

→ AnyLogic <https://www.anylogic.fr/road-traffic/>

AnyLogic fournit une Bibliothèque de trafic routier, qui offre la flexibilité et la puissance nécessaire pour parvenir à une conception du trafic routier plus efficace. Une visualisation claire est un moyen simple et rapide pour faciliter le développement de modèles d'infrastructures routières, avec des cartes de densité présentant la congestion et des animations montrant le flux du trafic et les goulets d'étranglement.

→ Citilog <https://www.citilog.com/>

Citilog crée une mobilité plus intelligente grâce à l'analyse vidéo et à l'intelligence artificielle appliquées aux solutions de gestion du trafic. Maîtrisant les derniers développements en Deep Learning pour l'analyse vidéo, Citilog permet un voyage plus fluide et plus sûr pour des millions de conducteurs à travers le monde. Ces solutions intelligentes réduisent les coûts et les risques pour les opérateurs de trafic, en fournissant des outils plus intelligents et en fin de compte une route plus sûre pour les voyageurs qu'ils desservent.

→ C-The Difference <https://www.neogls.com/innovation/c-the-difference/>

Vise à combler le vide qui sépare les implantations les plus avancées des STI-C dans un environnement urbain et les déploiements et opérations à plus grande échelle en ciblant les professionnels responsables des opérations et de la planification des transports urbains, les décideurs politiques et les décisionnaires finaux.

Ce groupe de pionniers croit fermement dans la capacité des services STI-C à apporter des solutions efficaces et efficientes pour résoudre des problèmes de mobilité urbaine tout en respectant la fluidité du trafic, la sécurité des usagers, mais également l'impact sur l'environnement.

#### **4: trafic routier**

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

Le débit et la vitesse/taux d'occupation.

Débit: mesurable par des dispositifs de comptage fixes.

Vitesse: représente la vitesse de déplacement du flot de véhicules

Le taux d'occupation représente le pourcentage de temps durant lequel un point de la route est « occupé » par un véhicule au-dessus de lui. Le taux d'occupation est exprimé en pourcentage. Il est à noter que sa qualité de mesure par des capteurs classiques est très dégradée lorsque les vitesses des véhicules sont basses.

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

Permet de calculer automatiquement un seuil d'encombrement pour chaque détecteur, à partir de la reconstruction du diagramme fondamental (tracé d'occupation en fonction du débit)

Les variables de débit, de vitesse et de concentration sont donc liées entre elles (par la relation  $Q=K*V$ , et la relation (dite fondamentale) qui les relie deux à deux constitue le diagramme fondamental (DF).

c- quel est le débit typique maximal d'un tronçon à une voie

L'unité UVP est l'unité de véhicule particulier définie comme suit :

un véhicule léger ou une camionnette = 1 UVP ;

un poids-lourds de 3,5 tonnes et plus = 2 UVP ;

On a donc les débits typiques maximaux suivants:

- en zone urbaine : 5 000 à 7 000 UVP/jour
- sur voie rapide ou autoroute : 30 000 UVP/jour
- Pourquoi cette différence? Sur les voies rapides ou autoroutes, les véhicules ne sont pas soumis aux effets des feux de signalisation, ou des croisements avec d'autres voies qui impliquent de ralentir. La route est souvent droite et le véhicule roulant dessus est souvent prioritaire aux véhicules qui entrent sur la voie.

## 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? (déterminées en cours)

- Origine (x, y)
- Destination (x, y)
- Distance (vol d'oiseau)
- Catégories de routes, nb voies
- Durée
- Durée statique
- Durée estimée avec les règles
- Heure
- Jour de la semaine
- Id mission

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

- *Expérience du conducteur*

- Motivation du conducteur (s'il doit se dépêcher...)
- Circonstances locales

## 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 si il n'existe pas d'algorithme polynomial pour le résoudre.

**FAUX**

*Un problème NP est un problème de décision vérifiant les propriétés suivantes :*

- *il est possible de vérifier une solution efficacement (en temps polynomial) ; la classe des problèmes vérifiant cette propriété est notée NP ;*
- *Tous les problèmes de la classe NP se ramènent à celui-ci via une réduction polynomiale ; cela signifie que le problème est au moins aussi difficile que tous les autres problèmes de la classe NP.*

*De plus, une propriété de la définition implique que s'il existe un algorithme polynomial pour résoudre un quelconque des problèmes NP-complets, alors tous les problèmes de la classe NP peuvent être résolus en temps polynomial.*

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

**VRAI**

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}  
allDifferent(u,v,w,x,y)

Faux 2 solutions u=1 v=2 w=3 Y=4 x=5 et u=3 v=2 w=4 x=1 y=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**

Le CP-Optimizer est un algorithme exact c'est-à-dire qu'il trouve une solution exacte et optimale pour un problème NP-difficile.

## 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.

Problème du sac à dos: Il modélise une situation de remplissage d'un sac à dos, ne pouvant supporter plus d'un certain poids. Avec à disposition un ensemble donné d'objets ayant chacun un poids et une valeur. Les objets mis dans le sac à dos doivent maximiser la valeur totale, sans dépasser le poids maximum.

Les objets sont numérotés par l'indice  $i$  variant de 1 à  $n$ . Les nombres  $w_i$  et  $p_i$  représentent respectivement le poids et la valeur de l'objet numéro  $i$ . La capacité du sac sera notée  $W$ .

Il existe de multiples façons de remplir le sac à dos. Pour décrire l'une d'elles, il faut indiquer pour chaque élément s'il est pris ou non. On peut utiliser un codage binaire : l'état du  $i$ -ème élément vaudra  $x_i = 1$  si l'élément est mis dans le sac, ou  $x_i = 0$  s'il est laissé de côté.

(sources :  
[https://fr.wikipedia.org/wiki/Probl%C3%A8me\\_du\\_sac\\_%C3%A0\\_dos#Complexit%C3%A9\\_et\\_cryptographie](https://fr.wikipedia.org/wiki/Probl%C3%A8me_du_sac_%C3%A0_dos#Complexit%C3%A9_et_cryptographie))

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

Quelles boîtes choisir afin de maximiser la somme emportée tout en ne dépassant pas les 15 kg autorisés ?

Les boîtes:

9kg - 10\$  
12kg - 7\$  
2kg - 1\$  
7kg - 3\$  
5kg - 2\$

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

Non optimale: algorithme glouton: ajouter en priorité les objets les plus efficaces, jusqu'à saturation du sac. total 11\$ pour 11kg.

Solution optimale: 12 \$ pour 14 kg

## 8. programmation par contraintes

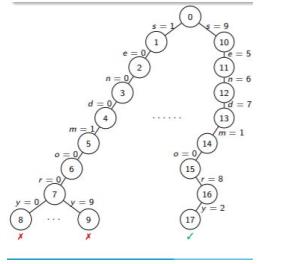
Deux principes fondamentaux de la Programmation par Contraintes sont (1) la recherche arborescente et (2) le filtrage du domaine des variables. Décrivez brièvement ces principes, leurs rôles et la façon dont ils sont mis en oeuvre durant la résolution.

(1) recherche d'arborescence:

Dans le cas de la résolution sur domaines finis, il est en théorie possible d'énumérer toutes les possibilités et de vérifier si elles violent ou non les contraintes. C'est la méthode Générer et tester

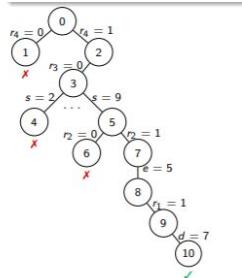
On génère toutes les affectations possibles et on vérifie si elles correspondent à des solutions.

Cependant, cela s'avère impraticable pour des problèmes de taille moyenne en raison du grand nombre de combinaisons possibles.



## (2) Filtrage

Une des principales parties de la résolution, appelée « filtrage », a pour but d'éviter cette énumération exhaustive. Elle consiste à déduire à partir des contraintes les valeurs impossibles. Lorsqu'une variable ne possède plus qu'un candidat, celle-ci est instanciée (i.e. cette valeur lui est affectée)



[https://fr.wikipedia.org/wiki/Programmation\\_par\\_contraintes](https://fr.wikipedia.org/wiki/Programmation_par_contraintes)