

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

FLEURY Orlanne | ENSC

1: IA

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

Dans une démarche scientifique basée sur une approche réaliste (ou platonicienne), la science a pour but de créer de la connaissance, d'explorer le « monde des idées » : l'objectif ici est la connaissance, sans visée applicative derrière. Au contraire, une approche utilitariste requiert de la science de servir à quelque chose : on doit pouvoir l'utiliser à des fins applicatives. On se soucie alors moins des concepts fondamentaux derrière.

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] et input = []

L'algorithme recherche le minimum de l'input et l'ajoute à result tout en le supprimant de l'input. Nous sommes donc ici sur un algorithme de tri

Sa complexité semble être quadratique

Lien utilisé : http://igm.univ-mly.fr/~pivotau/STRUCT/COURS/struct_cours2_1x4.pdf

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	Optimisation de tournée, organisation des tournées par zonage, tracking temps réel, GPS, prise en compte des spécificités, à partir des données du CRM, réorganisation des tournées simples https://www.mapotempo.com/mapotempo-web-logiciel-optimisation-tournees/
ERP Copilote (Infologic)	Organisation des tournées Prise en compte des différents type de commande Lié au transporteur Gestion des tournées en liaison avec des commandes (feuille de route, bon de livraison, encaissement...) Données temps réel Domaine agro-alimentaire https://www.youtube.com/channel/UCGaPfxIWnkTYP-oYxx2m5dg
AntsRoute	Plannification intelligente des tournées/interventions (optimisaiton des déplacements) Suivi agents temps réel Plannification avec intégration de contraintes Notifications clients Digitalisation processus en automatisant les rapports d'intervention https://antsroute.com/

4: trafic routier

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

ID du point de détecteur (elle ne se mesure pas mais est utilisée)

Date

Heure

Debit par heure du nombre de véhicule

Taux d'occupation

Ces variables sont ensuite utilisées pour déduire la densité du trafic

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 permet de représenter la congestion (le taux d'occupation) dans le temps. L'étude du diagramme peut alors faire ressortir des patterns utile à la prédiction de la congestion. Des seuils peuvent alors aussi apparaître

→ quand on augmente le nombre de véhicules, celui qui fait apparaître des ralentissements, lorsqu'il y a trop de véhicules et ils rouent trop près des uns des autres : le débit peut encore augmenter

→ quand ce nombre de véhicules augmente encore et qu'ils ne peuvent aller plus vite et le débit n'augmente pas – il est maximale.

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

- en zone urbaine
- sur voie rapide ou autoroute
- Pourquoi cette différence?

Le débit sur voie rapide/autoroute est supérieur à celui d'une zone urbaine du notamment aux infrastructures et aux différentes possibilités : en voie rapide les routes sont plus larges, sécurisées et dédiées à la vitesse (optimisation des entrées et des sorties pour garantir flux continu de voiture – on va toujours tout droit) alors qu'en zone urbaine les infrastructures sont plus complexes et moins fluides (virage, carrefour, feux tricolore)

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)

Distance du trajet

Identifiant de la tournée

Prévision du temps de trajet

Occupation de la chaussée

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

Circonstance locale comme des adresses incomplètes (manque information ou entrée à un numéro différent), du retard dans la livraison ou encore l'expérience du conducteur

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 de décision est dans la classe de complexité NP si il n'existe pas d'algorithme polynomial pour le résoudre ou s'il peut être vérifié par un vecteur X donné. Le Si et seulement si est donc faux

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

FAUX – La programmation par contrainte est plus performante

c- Le problème suivant possède exactement trois solutions:

$$u \in \{1,3\}$$

$$v \in \{1,2\}$$

$$w \in \{3,4\}$$

$$x \in \{1,5\}$$

$$y \in \{4,5\}$$

$$\text{allDifferent}(u,v,w,x,y)$$

FAUX – 2 solutions existent : (1;2;3;4;5) et (3;2;4;1;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

7. optimisation

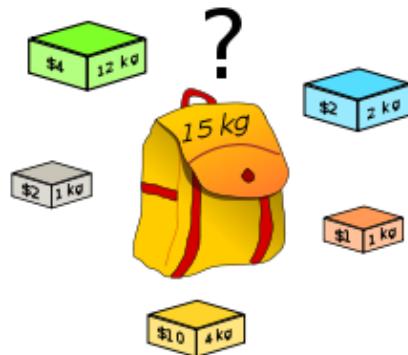
a- En cherchant sur internet, décrivez un problème d'optimisation combinatoire non vu dans le cours dont la version de décision est un problème NP-Complet.

Le problème du sac à dos : des objets définis par une valeur et une masse doivent être placés dans un sac à dos. Le but étant de maximiser la valeur totale des objets sans dépasser la masse maximale autorisée.

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

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

Une instance possible est celle proposée par Wikipedia sur la page du problème :



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

Une solution faisable et non optimale serait de prendre le cube vert, gris et orange, soit $4\$ + 2\$ + 1\$ = 1\$$, pour une masse de $12 + 1 + 1 = 14\text{kg}$

La solution optimale, après comparaison de l'ensemble des combinaisons, serait de prendre le cube gris, jaune, orange et bleu, pour un total de $15\$$ et 8kg .

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.

Dans la programmation par contrainte un problème est défini par un ensemble de variables de décision et une liste de contraintes logiques.

Dans la recherche arborescente, on est sur une approche constructive où l'on part du problème initial (le nœud racine) pour explorer un arbre en fixant une par une les variables de décision. A chaque branche, formant le premier niveau de sous arbre on va alors fixer pour une variable de décision toutes les valeurs possibles. Il faut ensuite commencer par une première branche et continuer à explorer tous ses sous arbres. Lorsqu'on arrive à une situation d'impossibilité on remonte en arrière pour tester les enfants du nœuds suivant. On teste alors l'ensemble des possibilités avant d'arriver à la solution, mais cela n'est pas un très optimal.

Le filtrage du domaine des variables consiste à enlever, pour chaque variable de décision, toutes les valeurs pour lequel il y a au moins une contrainte rendant ces valeurs impossibles. L'algorithme de filtrage est alors propre à chaque contrainte pour filtrer les variables.