

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

<< Vauchez Juliette >>

1: IA

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

On considère des approches pour aborder la connaissance (des mondes réel, des idées et des informations). L'approche réaliste a pour but de comprendre le monde à partir d'observations empiriques et de le formaliser. L'approche utilitariste quant à elle a pour but de servir à quelque chose, elle est plus appliquée. Elles ne prennent pas les éléments dans le même sens.

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.

On obtient `result = [0;2;4;5;9]` et `input = []`.

L'algorithme parcourt le tableau donné en entrée (`input`) et ajoute le minimum au tableau de résultat (`result`), de façon itérative jusqu'à ce que le tableau initial soit vide.

Il s'agit ici d'une complexité linéaire

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.

En cherchant sur internet avec les mots clés “logiciel commercial chemin de livraison optimal”, on trouve en premier Mapotempo (<https://www.mapotempo.com/>) qui est un logiciel open source déployable sur toutes les plateformes car basé sur une technologie web, qui permet une optimisation du temps de trajet en temps réel. Vient ensuite Portatour (<https://www.portatour.com/fr>) qui permet de planifier les trajets mais aussi le temps que l'on

souhaite passer avec les clients. Finalement, AntsRoute (<https://antsroute.com/>) permet d'augmenter la performance grâce à une optimisation des interventions par planification intelligente.

4: trafic routier

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

Le taux d'occupation, la vitesse, le débit (on en déduit alors la congestion par rapport au seuil de débit maximal admissible).

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

Le but est de déterminer les moments de congestion à partir de données historiques de la congestion sur chaque détecteur, on observe des courbes en cloche. Le diagramme fondamental est produit en mettant en abscisses le taux d'occupation (ou la vitesse) et en ordonnée le nombre de véhicules qui passent. On voit la courbe du taux d'occupation (c'est à dire de la congestion) en fonction des heures de la journée. On voit alors des patterns émerger, ce qui est utile pour prévoir la congestion. Lorsque le taux est faible, le débit augmente quand on augmente le nombre de véhicules, jusqu'à un certain seuil de 35% (le débit max admissible) où le débit diminue. Donc ce qui est en dessous du seuil est fluide, ce qui est dessus est encombré. C'est alors utile pour mesurer la congestion.

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

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

Parce qu'il y a plus de monde au même endroit sur une voie rapide qu'en zone urbaine il y a plus de virages et issues possibles. C'est donc "chaotique" dans le milieu urbain et difficile à prévoir car les relations entre les arcs et les graphs (tournants) sont difficiles à définir.

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)

La distance > le RoundID avec le type de tournée et de chauffeur (expérience) facteur endogène > temps trajet prévu > occupancy (pourcentage de temps où il y avait un véhicule sur la chaussée) liée à la vitesse > angle

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

Le taux d'erreur d'adresse fait que l'on est jamais sûr de la tournée du camion

Les 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 (car un problème de décision appartient à la classe P inclue dans NP si toutes ses instances peuvent être résolues avec un algorithme polynomial)

b- Dans l'industrie, la majorité des problèmes d'ordonnancement sont résolus grâce à des heuristiques. FAUX (La plupart sont des COP et on peut les résoudre avec de la programmation par contrainte)

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 (En réalisant un arbre, on trouve deux solutions (1,2,3,5,4) 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 (La recherche est complète de par l'algorithme exact, à n'importe quel temps, parallèle, aléatoire et déterministe)

7. optimisation

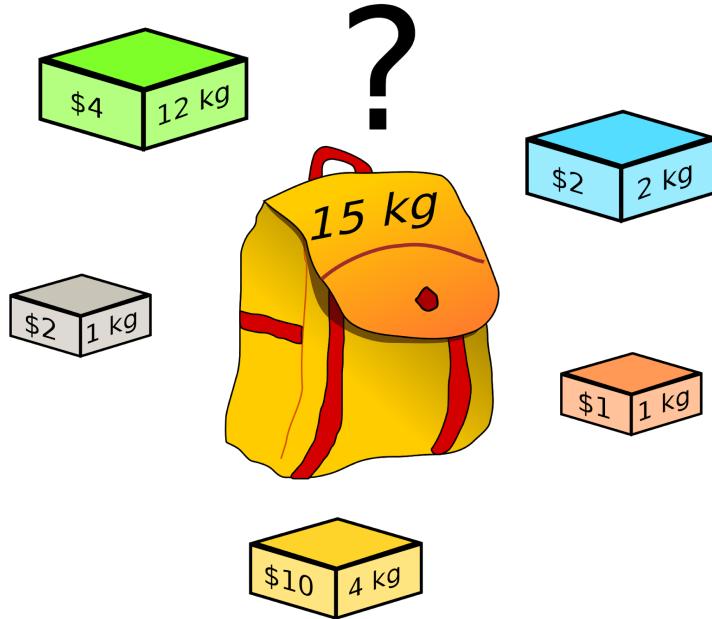
a- En cherchant sur internet, décrivez un problème d'optimisation combinatoire non vu dans le cours dont la version de décision est un problème NP-Complet.

On peut trouver le problème du sac à dos visant à remplir un sac à dos avec des objets qui ont une valeur et un poids donné sans dépasser le poids maximal autorisé par le sac. On veut emporter la valeur max :

(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 peut être la suivante, trouvée sur
https://fr.wikipedia.org/wiki/Probl%C3%A8me_du_sac_%C3%A0_dos :



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

On procède avec un arbre de décision incluant les alternatives de prendre ou non le poids : on calcule alors la somme des poids et la somme de leur valeur. On choisira comme meilleure solution la valeur et le poids maximums associés et nous pourrons mettre en avant d'autres solutions.

Une solution non-optimale (respectant quand même les conditions) serait un poids vert, un poids orange et un poids bleu, rapportant 7\$ pour 15kg.

La meilleure solution rapporte 15\$ pour 8kg avec un poids gris, un poids jaune et un poids orange et un poids bleu.

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 programmation par contrainte consiste à définir un ensemble de variables de décision ainsi qu'une liste de contraintes logiques (mathématiques). On peut définir ces deux principes fondamentaux de la façon suivante :

- Recherche arborescente : approche comportant un nœud racine avec toutes les variables initiales comme point de départ et proposant les variables accessibles à partir de chacun des nœuds suivants. On procède par propagation en explorant l'arbre et en essayant de fixer les variables de décision. Le but est de tester toutes les possibilités.
- Filtrage du domaine des variables : pour chaque contrainte, des algorithmes particuliers filtrent les variables qui interviennent. L'enjeu est d'être plus efficace en réduisant la combinatoire du problème (enlevant les variables non-pertinentes)

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