

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

Instructions:

- Copier/coller ce texte dans un mail adressé à baudelth@fr.ibm.com à la fin de l'examen, et remplir les réponses en dessous de chaque question. DEPUIS VOTRE ADRESSE ENSC.FR (un filtre sur le mail sera appliqué pour trier les réponses).
- **Le nom de votre fichier réponse doit commencer par votre nom de famille.prénom.**
- 5 à 6 lignes de texte par question sont généralement suffisantes pour obtenir une bonne note. Des points supplémentaires sont attribués pour des réponses plus détaillées. Chaque question apporte 2 points. Il n'est pas nécessaire de répondre aux questions marquées [BONUS] pour obtenir la note maximale mais des réponses justes à ces questions rapportent des points supplémentaires. Certaines questions sont beaucoup plus faciles que d'autres, et elles ne sont pas (toutes) par ordre de difficulté croissante...
- **Lorsque vous utilisez une réponse trouvée sur internet, donner l'hyperlien des sources utilisées.**

<<Cittadini, Maëlys>>

1: IA

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

https://sites.uclouvain.be/reso/opac_css/doc_num.php?explnum_id=17973

Une approche réaliste est une méthode qui, à partir d'une grande quantité d'information, permet de générer des données utiles pour la création et le développement d'un programme. A partir des données générées, on peut mettre en place un raisonnement ou déduire des comportements à adopter dans un système.

En revanche, une approche utilitariste repose sur le modèle de la pensée humaine mais sous forme informatique. On se rapproche de la psychologie humaine pour créer un nouveau programme.

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.

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

L'algorithme test toutes les valeurs qui sont supérieures à 0 et il prend la plus petite valeur qu'il rencontre et lui associe la variable `tmp`. On remarque que c'est une boucle qui ajoute plusieurs valeurs dans le dans la liste `result` et les supprime de la liste `input`.

Puis quand la variable `i` est inférieur à la valeur `tmp`, on lui ajoute `+1` et quand la variable `i` est supérieur a `tmp`, on ajoute `+1` à la variable `i`.

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.

Logiciel	SmartDeliveries	Danem / DUNDDL	Digiparc	AntsRoute
Site		https://danem.fr/index.php/fr/solutions-metiers/transport-livraison	https://digiparc.com/solutions/transport-tms.html	https://antsroute.com/
Optimisation des tournées	Oui	Oui	Oui	Oui
Tracking des livraison	Oui	Oui	Non	Oui
Prévision des tournées	Oui	Oui	Oui	Oui
Informer et prévoir le trafic	Oui	Oui /Waze	Non	Non
Information préalable sur la prévision de la tournée	Oui	Oui	Oui	Non
Les plus	Modification de l'itinéraire en temps réel basé et prise en compte des données enregistrées précédemment	Scan par code barre des colis : fonctionnement sans papiers	Gestion de la flotte, couts, maintenance et équipement.	Interaction avec les clients, analyse d la performance

4: trafic routier

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

Les deux variables sont le débit d'automobiles ainsi que la vitesse en fonction du taux d'occupation.

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 reprend la densité du trafic en fonction des jours mesurés. Il met en avant le trafic en fonction d'intervalles d'horaires et donc de prévenir la congestion le weekend ou sur des fourchettes d'heures. C'est un modèle robuste car mesuré du 1^{er} avril au 30 décembre 2011 de 6h du matin à 18h le soir. On remarque que le weekend du 19 novembre il y a eu moins de circulation qu'en semaine.

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

- en zone urbaine
155.08
- sur voie rapide ou autoroute
118.46
- Pourquoi cette différence?

Les livreurs ont plus souvent rendez-vous à des adresses personnelles, donc dans des lieux éloignés des grands axes. De plus, le GPS va offrir aux conducteurs la possibilité de faire des détours par des voies annexes quand certaines se retrouvent bouchées, cela va avoir une influence sur le trafic urbain. De ce fait, le débit sera plus important en zone urbaine qu'en voie rapide.

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)

L'importance relative a été évaluée et les critères les plus importants sont :

- Distance : 77
- Identification du livreur → Véhicule, expérience du conducteur, motivation : 11
- Prévision du temps de trajet : 6
- Congestion : 2
- Angle : 1

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

En TP, nous avons identifié les problèmes de renseignement de l'identification des colis ainsi que les accidents. On parle alors d'imprévus et du manque de précision résultant du temps des livreurs.

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

- 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

- 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.

<https://www.techno-science.net/glossaire-definition/Probleme-du-sac-a-dos.html>

Problème du sac à dos proposé par Richard Karp dans un article de 1972 :

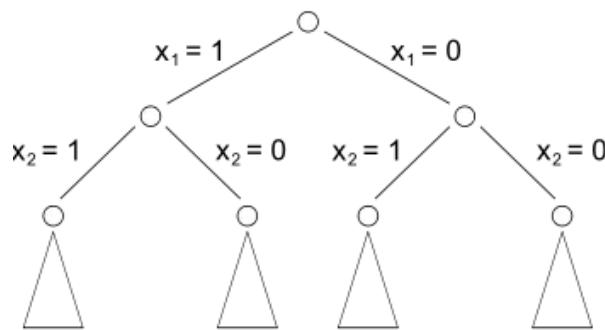
On à un sac à dos qui peut supporter un poids maximum, on veut le remplir de différents objets représentant à la fois une valeur mais aussi un poids. Tout le problème ici repose sur comment emporter la valeur maximum dans le sac sans dépasser le poids maximum ?

Pour la résolution de ce problème, 3 cryptographes ont mis au point un algorithme de chiffrement asymétrique.

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

La résolution façon optimisation est proche de la version décision. Ces deux versions proposent toutes deux une difficulté de résolution similaire.

Ainsi, dans la résolution de notre problème on observera cet arbre d'exploration :



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

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.

<https://tel.archives-ouvertes.fr/tel-00647479/document>

La recherche arborescente est un algorithme de recherche heuristique utilisé pour la prise de décision.

Elle se base sur une recherche de la meilleure solution sur des chemins déjà existants, déjà utilisés.

« Un nœud de l'arborescence correspond à une variable x du problème et ne branche issue de ce nœud représente l'instanciation de la variable x par une valeur de son domaine de telle sorte que cette instanciation soit cohérente avec les instanciations précédentes. »

Le filtrage du domaine des variables permet de supprimer des valeurs qui sont plusieurs fois contraintes par des modèles et ne pourront conduire à aucune solution de résolution.