

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

1: IA

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

Approche réaliste ce rapproche du but de la science, c'est à dire d'approfondir des connaissances dans le monde physique

L'approche utilitariste vis à extraire de l'information sans pouvoir approfondir la connaissance mais qui permet de répondre à un cas d'usage précis.

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

L'algorithme parcourt le tableau et cherche la plus petite valeur qu'il stocke dans la variable `tmp`, une fois arrivé à la fin du tableau il ajoute la valeur stockée dans `tmp` dans le tableau `result` et la retire du tableau `input`. C'est un algorithme de tri.

Cet algorithme est d'ordre N^2 .

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.

4: trafic routier

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

Densité, congestion

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

Graphique présentant en abscisse le taux d'occupation ou la vitesse et en ordonnées le nombre de voitures qui passent. Ce diagramme est utile pour prévoir la congestion car on peut y trouver des régularités en fonction des jours et des heures.

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

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

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, identifiant de la livraison, prévision du temps de parcours, occupation, 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.

L'identifiant de la livraison et les accidents

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.

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

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.

La recherche arborescente est une recherche heuristique utilisée dans le cadre de la prise de décision. L'algorithme dispose d'un arbre des possibles contenant des situations déjà explorées, une sélection parmi cet arbre est la base de cette recherche arborescente.

Le filtrage du domaine des variable permet de supprimer certaines valeurs de variables. Ces valeurs sont supprimées car on peut déterminer mathématiquement qu'elle ne prendront part à aucune situation gagnante, à cause de certaines contraintes.