Devoir surveillé – Intelligence artificielle : apprentissages, usages et biais

Apprendre à partir de données (entraînement, modèle, données de test, taux de réussite, score de confiance) ; les grands types d'apprentissage : supervisé, non supervisé, par renforcement et auto-supervisé ; les usages de l'IA (identification, traduction, géolocalisation, calcul) et sa place dans la chaîne d'information, faux positifs et faux négatifs ; les biais et les effets de l'IA sur la société

Durée : 55 min – Sans document – Sur 20 points · Technologie 3e

Nom : _____ Date : _____ Note : _____ / 20

Les réponses doivent être rédigées et justifiées. Soigne l'orthographe et la présentation.

Exercice 1 – Quatre IA, quatre apprentissages ★★ *(4 pts)*

Nommer le type d'apprentissage de chaque IA (supervisé, non supervisé, par renforcement ou auto-supervisé) et justifier à partir des données utilisées.

a) Un logiciel regroupe des milliers de morceaux de musique en « ambiances » d'après leur tempo et leurs instruments, sans liste d'ambiances fournie.

b) Avant d'apprendre à traduire, un modèle apprend à retrouver des mots effacés dans des milliards de phrases.

c) Un bras robotisé gagne des points à chaque objet saisi et en perd à chaque objet lâché.

d) Un logiciel apprend à repérer les grains de café abîmés sur 30 000 photos triées « abîmé » ou « sain » par des experts.

Exercice 2 – La mangeoire connectée ★★ *(4 pts)*

Une mangeoire photographie chaque oiseau qui se pose ; les photos partent par Wi-Fi vers un serveur où un modèle reconnaît l'espèce, qu'une application affiche sur le téléphone. Le modèle a appris (apprentissage supervisé) sur 12 000 photos étiquetées, dont 10 % gardées pour le test.

a) Calculer le nombre de photos d'entraînement et de test, puis le taux de réussite si 1 146 photos de test sont bien reconnues.

b) Donner la fonction de la chaîne d'information réalisée par la caméra, par le modèle et par l'application. Le modèle est-il embarqué ? Que se passe-t-il si Internet est coupé ?

c) L'espèce n'est affichée que si le score de confiance atteint au moins 70 %. Scores d'un matin : 91 %, 70 %, 55 %, 83 %, 69 %, 97 %. Combien de réponses sont affichées ? Un élève affirme : « 97 %, c'est forcément une mésange charbonnière. » Commenter.

d) L'application prévient aussi quand un chat approche (positif = « alerte chat »). Un écureuil déclenche l'alerte ; un chat passe la nuit sans alerte. Nommer ces deux erreurs ; laquelle est la plus grave pour les oiseaux ?

Exercice 3 – Les plus proches voisins d'un vélo électrique ★★★ *(4 pts)*

Un vélo électrique reconnaît le type de voie avec deux mesures : un indice de vibrations (de 0 à 10) et la vitesse (en km/h). Exemples étiquetés « route » : R1 (2 ; 24), R2 (3 ; 21), R3 (1 ; 18) ; étiquetés « chemin » : C1 (7 ; 14), C2 (8 ; 11), C3 (6 ; 17). Nouvelle mesure : N (4 ; 18). Distance entre deux mesures, comme l'hypoténuse d'un triangle rectangle : d = $\sqrt{(\text{écart des vibrations}^2 + \text{écart des vitesses}^2)}$.

a) Calculer la distance entre N et C3, puis entre N et R3 (au centième si besoin).

b) Les autres distances valent : N–R1 ≈ 6,32 ; N–R2 ≈ 3,16 ; N–C1 = 5 ; N–C2 ≈ 8,06. Ranger les six exemples du plus proche au plus éloigné de N, puis donner la réponse du modèle avec 1 voisin.

c) Donner la réponse du modèle avec 3 voisins, puis avec 5 voisins.

d) C3 a en fait été mesuré sur une piste cyclable lisse et étiqueté « chemin » par erreur. Nommer cette cause d'erreur. Pourquoi le choix de 3 voisins en a-t-il limité l'effet ? Pourquoi prend-on un nombre impair de voisins ?

Exercice 4 – Deux itinéraires pour aller au stade ★★★ *(4 pts)*

Pour aller du collège au stade, une application de navigation compare deux itinéraires. Son modèle prévoit la vitesse sur chaque tronçon à partir de l'historique du trafic et des positions envoyées par les téléphones des usagers. Itinéraire A : 6 km à 30 km/h, puis 4 km à 60 km/h. Itinéraire B : 9 km à 54 km/h, puis 3 km à 36 km/h (vitesses prévues).

a) Nommer l'usage de l'IA mis en jeu et la fonction de la chaîne d'information réalisée par le GPS des téléphones. Pourquoi les positions envoyées sont-elles des données personnelles ?

b) Calculer en minutes la durée prévue de chaque itinéraire (durée = distance ÷ vitesse). Lequel l'application propose-t-elle ?

c) Sur le 2e tronçon de B, un camion de livraison ralentit tout le monde : la vitesse réelle est de 18 km/h. Calculer la durée réelle de B et l'écart avec la prévision. A, s'il était bien prévu, aurait-il été plus rapide ?

d) Pour éviter un bouchon, l'application envoie des centaines de conducteurs dans la même petite rue résidentielle. Citer deux effets de cet usage sur la société et proposer un moyen de les limiter.

Exercice 5 – Une enceinte qui comprend mal les enfants ★★★ *(4 pts)*

Le modèle de reconnaissance vocale d'une enceinte connectée a appris sur 50 000 heures d'enregistrements, dont 45 000 heures de voix d'adultes et le reste de voix d'enfants. Test : 500 commandes d'adultes, dont 10 mal comprises ; 250 commandes d'enfants, dont 35 mal comprises. La publicité annonce : « 94 % des commandes comprises ! »

a) Calculer la part des voix d'enfants dans les données d'entraînement.

b) Calculer le taux d'erreur des adultes et celui des enfants. Combien de fois plus d'erreurs les enfants subissent-ils ?

c) Vérifier le chiffre de la publicité, puis expliquer pourquoi il est trompeur.

d) Nommer la cause de ce biais et proposer deux corrections. Citer un autre effet de cette enceinte sur la vie privée ou sur l'environnement.

Bon courage !

Ce document est la propriété intellectuelle de Ben, professeur particulier (benprofparticulier.fr), protégée par le Code de la propriété intellectuelle (art. L. 111-1 et L. 122-4). Il est mis à disposition pour un usage personnel de révision. Toute reproduction, diffusion, mise en ligne ou revente, totale ou partielle, sans autorisation écrite est interdite et constitue une contrefaçon (art. L. 335-2).