

Projet d'Initiation à la Recherche ou ~~Projet d'Innovation-Recherche~~ (barrer)

Nom du laboratoire ou de l'entreprise/établissement :

LAAS-CNRS

TUTEUR(S)

NOM-Prénom: Nawal GUERMOUCHE Tel

Mel : guermouc@insa-

toulouse.fr

Tuteurs INSA si projet industriel:

Prénom NOM

Mel:

TITRE DU PROJET

Augmentation et génération de données intelligentes pour les algorithmes d'IA

MOT-CLES

IA, deep learning, generative IA, génération de données, augmentation des données

DESCRIPTIF (RESUME), indiquer l'enjeu sociétal de l'INSA de Toulouse s'il y a lieu

Contexte

Les systèmes de mobilité intelligente s'appuient de plus en plus sur des techniques d'intelligence artificielle afin de mieux comprendre l'état du conducteur, anticiper les situations à risque et proposer des mécanismes d'assistance adaptés. Dans le contexte de la mobilité inclusive, ces systèmes doivent également prendre en compte les besoins spécifiques de certains usagers, notamment les personnes présentant un trouble du spectre de l'autisme (TSA).

Pour ces usagers, certaines situations de conduite ou de déplacement, trafic dense, bruit, interactions complexes, changements imprévus, environnement inhabituel, peuvent entraîner une augmentation de la charge cognitive ou du stress. Concevoir des systèmes capables d'anticiper ces situations nécessite de disposer de données décrivant simultanément le contexte de mobilité, les événements rencontrés et éventuellement les réponses physiologiques ou comportementales de l'utilisateur.

Cependant, la collecte de telles données est coûteuse et complexe. Elle nécessite des expérimentations avec des participants, produit souvent des jeux de données de taille limitée et soulève également des questions de confidentialité et de protection des données personnelles.

La génération de données synthétiques constitue une piste intéressante pour compléter les données réelles. Elle consiste à produire artificiellement de nouvelles observations présentant des propriétés statistiques et temporelles similaires aux données originales. Ces données peuvent notamment servir à

augmenter les jeux d'apprentissage, simuler des situations rares ou difficiles à observer et faciliter l'expérimentation de nouveaux modèles d'intelligence artificielle.

Objectifs

Ce projet a pour but d'étudier et d'expérimenter différentes méthodes d'IA pour la génération de données appliquées à un scénario de mobilité inclusive pour des personnes avec TSA.

Plus précisément, les étudiants devront :

- comprendre les enjeux associés à la collecte de données dans les systèmes de mobilité inclusive
- étudier les principales familles de méthodes de génération de données synthétiques
- sélectionner et mettre en œuvre une ou plusieurs méthodes adaptées aux données considérées
- générer un jeu de données synthétiques à partir d'un jeu de données réel ou d'un sous-ensemble de celui-ci
- analyser la similarité entre les données originales et les données générées
- étudier dans quelle mesure les données synthétiques préservent les relations entre les différentes variables
- évaluer leur intérêt pour une tâche simple d'apprentissage automatique, par exemple la classification d'un niveau de stress ou la prédiction d'un indicateur physiologique.

À l'issue du projet, les étudiants devront fournir :

- une synthèse bibliographique sur la génération de données synthétiques
- une analyse exploratoire du jeu de données étudié
- l'implémentation d'au moins une méthode de génération de données
- un jeu de données synthétiques
- une comparaison quantitative et visuelle entre données réelles et synthétiques
- une première évaluation de l'intérêt de ces données pour une tâche de machine learning

PROFIL DES ETUDIANTS SOUHAITE (1 seul choix par projet)

- ☐ AE-SE : spécialité Automatique-Electronique parcours Systèmes Embarqués
- ☒ IR-SI : spécialité Informatique parcours Systèmes Informatiques
- ☒ IR-SC : spécialité Informatique parcours Systèmes Communicants

☐ (optionnel) ce projet peut être proposé à un ou des étudiants d'échange sur la partie réalisation seule (semestre 1 et/ou semestre 2)

PRIORITE : si vous posez plusieurs sujets, indiquer ici la priorité de ce sujet (1= plus prioritaire) : 1