OFFRE DE STAGE

ARD4PL : Vers la formalisation de la fonction de récompense adaptative pour apprendre des préférences des utilisateurs au sein des systèmes de transport multimodaux

Informations sur l'offre de stage

Encadrement	Kamaldeep Singh Oberoi (CESI, CESI-LINEACT), Guy Camilleri (Université de Toulouse, IRIT)
Sujet de l'offre de stage	Vers la formalisation de la fonction de récompense adaptative pour apprendre des préférences des utilisateurs au sein des systèmes de transport multimodaux
Dates du stage	5 à 6 mois - à partir de Février/Mars 2027 selon la disponibilité du/de la candidat.e
Lieu du stage	IRIT, Campus Rangueil, Université de Toulouse
Modalités de candidature	Les candidatures sont à adresser par mail à ksoberoi@cesi.fr et guy.camilleri@irit.fr avec votre CV à jour, une lettre de motivation et votre relevé de notes

La mission du stage

Contexte

Au cours des dernières années, l'intérêt croissant pour la promotion de la mobilité urbaine durable a conduit à l'étude de diverses approches visant à inciter les usagers à privilégier des modes de transport durables pour leurs déplacements quotidiens. L'une de ces approches repose sur la « personnalisation », qui vise à faire évoluer les comportements de mobilité en recommandant des itinéraires « personnalisés » correspondant étroitement aux préférences et aux besoins de l'usager [1, 2]. Récemment, outre l'utilisation de modèles fondés sur l'apprentissage profond pour la prédiction et la personnalisation de la mobilité [3], les algorithmes d'apprentissage par renforcement (RL) sont devenus de plus en plus utilisés pour prédire la prochaine localisation [4] et le prochain trajet [5], tout en apprenant les préférences des usagers [6], avec pour objectif ultime de personnaliser les services de mobilité.

Les algorithmes fondés sur RL formalisent le problème de l'apprentissage des préférences comme un problème de prise de décision séquentielle, à l'aide d'un processus de décision markovien (MDP), composé de l'espace des états, de l'espace des actions, de la probabilité de transition, de la fonction de récompense et du facteur d'actualisation [4, 5, 6]. En RL, la fonction de récompense joue un rôle très essentiel dans l'orientation du processus d'apprentissage de l'agent au sein d'un environnement non ergodique [7]. Les approches existantes qui appliquent le RL à la prédiction de la mobilité et à l'apprentissage des préférences des usagers tendent à conserver une fonction de récompense statique, c'est-à-dire qu'elle n'évolue plus une fois que l'agent a convergé vers la politique optimale. Cette approche fonctionne bien pour l'apprentissage des préférences à court terme.

Cependant, il a été démontré que les préférences de mobilité des usagers peuvent évoluer tout au long de leur vie (un concept connu sous le nom de *Mobility Biography* [8]). Par ailleurs, les modifications ponctuelles des infrastructures de transport peuvent également avoir une incidence sur les choix modaux des usagers [9].

Objectifs

À travers ce stage, nous souhaitons étudier l'applicabilité des algorithmes fondés sur RL à l'apprentissage des préférences de mobilité sur le long terme. Autrement dit, l'objectif est de combiner le RL avec des concepts issus de l'apprentissage continu (*Continual Learning*, CL), afin de développer des agents capables d'adapter leurs fonctions de récompense et de converger vers des politiques optimales pour l'apprentissage des préférences tout au long de la vie, en tenant compte des évolutions des infrastructures de transport (par exemple, le prolongement d'une ligne de métro existante ou la création d'une nouvelle ligne).

Le/La stagiaire travaillera sur la formalisation théorique du problème d'apprentissage des préférences tout au long de la vie sous la forme d'un processus de décision markovien (MDP). Il/Elle devra proposer une approche de formalisation de la fonction de récompense adaptative (*Adaptive Reward Design*, ARD) [10] permettant de guider le processus d'apprentissage au fil du temps, tout en préservant les connaissances précédemment acquises et en évitant l'oubli catastrophique (*catastrophic forgetting*) [11]. Il/Elle s'appuiera également sur des données de mobilité ouvertes et synthétiques afin de proposer des scénarios d'évaluation à différentes échelles spatio-temporelles. L'utilisation de simulateurs de mobilité reconnus, tels que SUMO (https://eclipse.dev/sumo/), MATSim (https://matsim.org/) ou encore GAMA (https://gama-platform.org/), pourra être envisagée pour générer des données synthétiques et/ou entraîner l'agent.

Missions

Ce stage a pour objectifs de :

1. Appliquer la formalisation de la fonction de récompense adaptative (*Adaptive Reward Design*, ARD) à l'apprentissage des préférences de mobilité sur le long terme.
2. Réaliser une brève revue de la littérature sur les approches fondées sur l'apprentissage par renforcement (RL) et l'apprentissage continu (CL) pour l'apprentissage des préférences des usagers.
3. Proposer une formalisation théorique d'un processus de décision markovien (MDP) intégrant l'ARD.
4. Étudier le rôle de l'ARD et de l'oubli catastrophique (*catastrophic forgetting*) dans l'apprentissage des préférences sur le long terme.
5. Évaluer l'approche à travers des scénarios d'évaluation multi-échelles, combinant le suivi de l'ensemble du parcours de vie d'un usager individuel (échelle microscopique) et l'étude de l'impact des évolutions des infrastructures de transport sur les choix modaux à l'échelle macroscopique.

Profil recherché

Le candidat doit :

1. Être inscrit en deuxième année de master (M2) en mathématiques appliquées ou en informatique, et disposer de solides compétences en Python.
2. Avoir une première expérience des algorithmes d'apprentissage par renforcement (RL), ainsi qu'un regard critique sur leurs avantages et leurs limites théoriques serait un plus. Des connaissances dans le domaine de l'apprentissage continu (*Continual Learning*, CL) constitueraient aussi un atout.
3. Manifester un intérêt pour la formalisation théorique de l'intelligence artificielle, au-delà de la simple implémentation des algorithmes.
4. Être particulièrement intéressé par la mobilité intelligente en tant que domaine d'application.
5. Faire preuve d'autonomie, de curiosité et d'un intérêt marqué pour la recherche.
6. Être capable de lire des articles en anglais.

Références

1. Oberoi, K.S., 2024. "Personalisation in Mobility-as-a-Service: Where we are and how to move forward". In the 10th International Conference on Vehicle Technology and Intelligent Transport Systems (VEHITS 2024) (pp. 352-360).
2. van Ardenne, M.T., Cebecauer, M., Cats, O. and Ma, Z., 2025. "Personalised passenger information systems in public transport: a review and a 5-level personalisation taxonomy". *Transport Reviews*, *45* (6), pp.1016-1047.
3. Luca, M., Barlacchi, G., Lepri, B. and Pappalardo, L., 2021. "A survey on deep learning for human mobility". *ACM Computing Surveys (CSUR)*, *55* (1), pp.1-44.
4. Luo, Z. and Miao, C., 2022, February. "RLMob: Deep reinforcement learning for successive mobility prediction". In *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining* (pp. 648-656).
5. Kermali, A., Oberoi, K.S., El Khalfi, Z. and Dupuis, Y., 2026. "Context-aware hybrid reinforcement and recurrent learning for student home-campus trip prediction". *Transportmetrica B: Transport Dynamics*, *14* (1), p.2693929.
6. Chu, K.F. and Guo, W., 2023. "Deep reinforcement learning of passenger behavior in multimodal journey planning with proportional fairness". *Neural Computing and Applications*, *35* (27), pp.20221-20240.
7. Baumann, D., Noorani, E., Mustafin, A., Sheng, X., Verbruggen, B., Vanhoyweghen, A., Ginis, V. and Schön, T.B., 2026. "Ergodicity in reinforcement learning". *arXiv preprint arXiv:2603.10895*.
8. Chatterjee, K. and Scheiner, J., 2015. "Understanding changing travel behaviour over the life course: Contributions from biographical research."
9. Brands, T., Dixit, M. and van Oort, N., 2020. "Impact of a new metro line in Amsterdam on ridership, travel times, reliability and societal costs and benefits". *European Journal of Transport and Infrastructure Research*, *20* (4), pp.335-353.
10. Baybas, R., D'Eramo, C. and Brune, P., "Adaptive Reward Design in Reinforcement Learning: A Taxonomy and Survey".
11. Van de Ven, G.M., Soures, N. and Kudithipudi, D., 2024. "Continual learning and catastrophic forgetting". *arXiv preprint arXiv:2403.05175*.

Financement du stage – la fédération MIDOC

Le stage est financé par le Défi Clé Mobilité Intelligente et Durable en OCcitanie. Le Défi Clé MIDOC, initié et soutenu par la Région Occitanie, est porté par l'Université de Toulouse et lie 16 établissements partenaires. 26 laboratoires et centres de recherche occitans sont impliqués au sein de la Fédération MIDOC qui vise à associer plusieurs champs de recherche autour du véhicule autonome, acceptable et connecté et des services de mobilités durables et centrés utilisateurs. Pour en savoir plus : https://midoc.univ-toulouse.fr/

INTERNSHIP OFFER

ARD4PL: Towards adaptive reward design for reinforcement-learning based user preference learning in multimodal transport systems

Information about the internship offer

Supervision	Kamaldeep Singh Oberoi (CESI, CESI-LINEACT), Guy Camilleri (Université de Toulouse, IRIT)
Topic	Towards adaptive reward design for reinforcement-learning based user preference learning in multimodal transport systems
Dates	5 to 6 months - starting from Feb/March 2027 according to the candidate's availability
Location	IRIT, Campus Rangueil, Université de Toulouse
Application	Applications should be sent by email to ksoberoi@cesi.fr and guy.camilleri@irit.fr with your CV, a motivation letter and recent mark sheets

Description

Context

Over the past few years, due to the growing interest for promoting sustainable urban mobility, various approaches have been studied to incite users to choose sustainable modes of transport for their daily mobility. One such approach is personalisation, which aims to change people's travel behaviour by recommending "personalised" itineraries closely matching the user's preferences and needs [1, 2]. Recently, in addition to applying deep learning-based models for mobility prediction and personalisation [3], reinforcement learning-based (RL) algorithms have become prevalent for predicting next location [4] and next trip [5], while learning user preferences [6], with the ultimate objective of personalizing mobility services.

RL-based algorithms formulate the problem of preference learning as a sequential decision-making problem using a Markov Decision Process (MDP) consisting generally of the state space, the action space, the transition probability, the reward function and the discount factor [4, 5, 6]. Typically, in RL, the reward function plays a key role in guiding the agent's learning process in a non-ergodic environment [7]. Existing approaches which apply RL for mobility prediction and learning user preferences tend to keep their reward function static, i.e., it does not change once the agent has converged to the optimal policy. This design choice works well for learning short-term preferences. However, it has been shown that user mobility preferences could vary over the lifetime of the user (a concept known as Mobility Biography [8]), and changes in occasional transport infrastructure impact users' mode choices [9] as well.

Objectives

With this internship, we aim to study the applicability of RL-based algorithms for long-horizon mobility preference learning. In other words, the objective is to combine RL with Continual Learning-based (CL) concepts so as to develop agents capable of updating their reward functions to converge on optimal policies for life-long preference learning taking into account the changes in the transport infrastructure (like extending an existing metro line or adding a new metro line).

The intern would work on the theoretical formalisation of the life-long preference learning problem as an MDP, propose adaptive reward design (ARD) [10] for guiding the learning process over time while retaining previous knowledge and avoiding catastrophic forgetting [11], and use open and synthetic mobility data to propose multi-scale evaluation scenarios. The use of well-known mobility simulators like SUMO (https://eclipse.dev/sumo/), MATSim (https://matsim.org/) and even GAMA (https://gama-platform.org/) could be envisaged for generating synthetic data and/or training the agent.

Missions

This internship aims to:

1. Apply the adaptive reward design (ARD) for long-horizon mobility preference learning
2. Conduct a brief literature review about RL and CL based approaches for learning user preferences
3. Propose a theoretical formalisation of MDP integrating ARD
4. Study the role of ARD and catastrophic forgetting for long-horizon preference learning
5. Evaluate the approach with multi-scale scenarios combining complete lifetime of an individual user (microscopic level) as well as the impact of changes to the transport infrastructure on macroscopic mode choices.

Desired candidate profile

The candidate must:

1. Be enrolled in an M2 applied mathematics or computer science programme, with solid experience in Python
2. Have some prior experience of reinforcement learning algorithms. A critical view of their theoretical advantages and disadvantages would be appreciated. Knowledge about the field of Continual Learning would be a plus
3. Demonstrate an interest in the theoretical foundations and formalisation of artificial intelligence, beyond simply implementing algorithms.
4. Be passionate about intelligent mobility as a field of application
5. Be independent and curious about research
6. Have a good level of English and French (to be able to read scientific papers in English)

References

1. Oberoi, K.S., 2024. "Personalisation in Mobility-as-a-Service: Where we are and how to move forward". In the 10th International Conference on Vehicle Technology and Intelligent Transport Systems (VEHITS 2024) (pp. 352-360).
2. van Ardenne, M.T., Cebecauer, M., Cats, O. and Ma, Z., 2025. "Personalised passenger information systems in public transport: a review and a 5-level personalisation taxonomy". *Transport Reviews*, *45* (6), pp.1016-1047.
3. Luca, M., Barlacchi, G., Lepri, B. and Pappalardo, L., 2021. "A survey on deep learning for human mobility". *ACM Computing Surveys (CSUR)*, *55* (1), pp.1-44.
4. Luo, Z. and Miao, C., 2022, February. "RLMob: Deep reinforcement learning for successive mobility prediction". In *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining* (pp. 648-656).
5. Kermali, A., Oberoi, K.S., El Khalfi, Z. and Dupuis, Y., 2026. "Context-aware hybrid reinforcement and recurrent learning for student home-campus trip prediction". *Transportmetrica B: Transport Dynamics*, *14* (1), p.2693929.
6. Chu, K.F. and Guo, W., 2023. "Deep reinforcement learning of passenger behavior in multimodal journey planning with proportional fairness". *Neural Computing and Applications*, *35* (27), pp.20221-20240.

7. Baumann, D., Noorani, E., Mustafin, A., Sheng, X., Verbruggen, B., Vanhoyweghen, A., Ginis, V. and Schön, T.B., 2026. "Ergodicity in reinforcement learning". *arXiv preprint arXiv:2603.10895*.
8. Chatterjee, K. and Scheiner, J., 2015. "Understanding changing travel behaviour over the life course: Contributions from biographical research."
9. Brands, T., Dixit, M. and van Oort, N., 2020. "Impact of a new metro line in Amsterdam on ridership, travel times, reliability and societal costs and benefits". *European Journal of Transport and Infrastructure Research*, *20* (4), pp.335-353.
10. Baybas, R., D'Eramo, C. and Brune, P., "Adaptive Reward Design in Reinforcement Learning: A Taxonomy and Survey".
11. Van de Ven, G.M., Soures, N. and Kudithipudi, D., 2024. "Continual learning and catastrophic forgetting". *arXiv preprint arXiv:2403.05175*.

Internship funding – MIDOC Federation

The internship is funded by the Défi Clé Mobilité Intelligente et Durable en OCcitanie. The MIDOC Key Challenge, initiated and supported by the Occitanie Region, is led by the University of Toulouse and involves 16 partner institutions. 26 laboratories and research centers in the Occitanie region are involved in the MIDOC federation, which aims to combine several fields of research around autonomous, acceptable, and connected vehicles and sustainable, user-centred mobility services. For more information: https://midoc.univ-toulouse.fr.