Sujet de stage de master
Détection sécurisée de dérives comportementales à long terme chez les personnes âgées via une approche hybride d'apprentissage

N. Bennani[2] and M. Thilliez[1]

[1]LAMIH – Département Informatique à completer marie.thilliez@uphf.fr
[2] Laboratoire d'InfoRmatique en Image et Systèmes d'information, UMR 5205 CNRS , INSA Lyon, France. nadia.bennani@insa-lyon.fr

1 Contexte

Le vieillissement de la population observé aujourd'hui fait émerger de nouveaux enjeux, notamment en matière de maintien à domicile des personnes âgées, afin de favoriser leur autonomie et de leur permettre de vieillir dans les meilleures conditions possibles. Les technologies de smart home, reposant sur des capteurs environnementaux (mouvement, ouverture de portes, utilisation d'objets), permettent de collecter des données sur les activités de la vie quotidienne. Ces données offrent un potentiel important pour analyser les comportements des personnes et détecter des signes précoces de déclin cognitif afin d'anticiper et de prévenir au mieux ces problèmes.

Les travaux existants, dans ce domaine, se concentrent principalement sur la reconnaissance d'activités [4, 3] ou la détection d'événements ponctuels tels que la chute [2] par exemple. Cependant, ces approches sont souvent insuffisantes pour identifier des changements progressifs et durables du comportement, pouvant être des indicateurs précoces de déclin cognitif, et/ou de perte d'autonomie, de l'apparition de maladies neuro-dégénératives telles que la maladie d'Alzheimer.

Les premiers travaux [1] sur l'analyse du comportement dans les smart homes se sont concentrés sur la modélisation des routines quotidiennes et sur l'observation, chez le patient, des changements de comportement sur de longues périodes. Ces travaux reposent sur des architectures d'apprentissage centralisées, dans lesquelles les données sont collectées, traitées puis utilisées dans le processus d'apprentissage.
Deux architectures sont alors envisagées : la collecte et l'apprentissage sur les données se font sur un serveur distant, ce qui soulève des questions de confidentialité et complique de fait l'acceptabilité de la solution ; ou la collecte et l'apprentissage se font localement chez le patient mais nécessitent une infrastructure performante pour traiter la masse de données collectées sur le long terme. Ainsi, proposer une approche hybride d'apprentissage permettrait d'aborder les problématiques de confidentialité des données et de performances.

Par ailleurs, la centralisation des données personnelles soulève des enjeux majeurs de confidentialité et de ce fait d'acceptabilité par les personnes âgées d'un suivi qui pourrait porter atteinte à leur vie privée. Les premiers travaux sur l'analyse du comportement dans les smart homes se sont concentrés sur la modélisation des routines quotidiennes. L'objectif est de construire un profil comportemental de référence à partir des activités observées. Dans cette perspective, les travaux présentés dans l'article [1] proposent une approche basée sur l'apprentissage automatique pour observer les changements de comportement chez les personnes âgées sur de longues périodes. La méthode repose sur la modélisation des ADL et sur la comparaison de profils comportementaux à l'aide de mesures de similarité temporelle. Cette approche permet de détecter des variations progressives du comportement, contrairement aux méthodes focalisées sur des anomalies ponctuelles. Elle met en évidence l'interêt d'observer et d'interpréter les changements comportementaux comme des indicateurs précoces de dégradation de l'état de santé. Ces travaux reposent sur des architectures centralisées, dans lesquelles les données sont collectées et traitées sur un serveur unique, ce qui peut soulever des questions de confidentialité et de risques potentiels d'inférences sur les données et ainsi compliquer l'acceptabilité de la solution. Face à ces enjeux, l'apprentissage fédéré s'impose comme une solution prometteuse. Ce paradigme permet d'entraîner des modèles localement en ne transmettant que les propriétés des modèles au serveur central qui agrège et affine le modèle global. Un tel paradigme préserve la confidentialité des données brutes et allège les transferts de données entre clients et serveur, tout en bénéficiant de l'apport des autres modèles pour une reconnaissance plus précises des activités humaines. Dans l'article [2], les auteurs proposent un cadre d'apprentissage fédéré pour le suivi des activités en temps réel à partir de capteurs

embarqués. Leur approche montre qu'il est possible d'obtenir des performances comparables à celles d'un modèle centralisé tout en préservant la confidentialité des données des utilisateurs.Dans le domaine des smart homes, les travaux [3] introduisent une approche basée sur des modèles LSTM et l'apprentissage fédéré pour la reconnaissance d'activités dans des environnements multi-utilisateurs. Ces travaux mettent en évidence les défis liés à l'hétérogénéité des données, caractéristique essentielle des environnements réels. Dans [5], les auteurs proposent FedMeet une approche intégrant une fusion adaptative des capteurs, une modélisation temporelle basée sur BiLSTM, une personnalisation au niveau des clients et un mécanisme de masquage des mises à jour des modèles qui empêchent des attaques par reconstruction renforçant ainsi la protection de la vie privée. Malgré ces avancées, les applications de l'apprentissage fédéré restent principalement limitées à la reconnaissance d'activités ou à la détection d'anomalies ponctuelles. L'intégration de ce paradigme dans des approches de suivi longitudinal du comportement demeure encore peu explorée. L'article [4] propose une évolution du paradigme de suivi des personnes âgées. Les auteurs soulignent la nécessité de passer d'une logique de détection d'événements isolés à une compréhension globale et continue du comportement. Cette vision met en avant l'importance : de l'analyse à long terme, de la contextualisation des activités et du choix de solutions respectueuses de la vie privée L'apprentissage fédéré dans ce contexte se confronte aux challenges suivants : (1) performance de l'apprentissage : l'apprentissage sur le long terme nécessite de traiter les données brutes sur une longue période avec donc une très grande quantité des données et donc un besoin important en ressources. (2) un probleme de confidentialité : en effet, les modèles transmis par les personnes atteintes de dérive du comportement seront faciles à repérer et peuvent ainsi révéler des informations confidentielles sur les personnes concernées.

C'est dans ce contexte que s'inscrit ce sujet de stage, dont l'objectif est de revisiter les techniques d'apprentissage fédéré pour les adapter aux besoins de la détection de dérives comportementales à long terme dans le suivi des personnes âgées.

2 Etapes du stage

Dans ce sujet, nous souhaitons étudier l'apport d'une approche hybride d'apprentissage basée sur le split learning pour la détection de dérives comportementales à long terme. Les étapes du projet sont définies comme suit :

- Identification des données et modélisation de comportements à long terme : il existe dans la littérature différents travaux qui présentent des solutions pour la représentation des habitudes de vie. Il sera nécessaire dans un 1er temps, d'identifier quels sont les types de données que nous souhaitons considérer et sur quels jeux de données nous allons travailler.
- Implémentation d'un apprentissage centralisé de type deep learning sur ces données pour détecter les dérives de comportement.
- Etude du split learning et définition des différents modèles de partitionnement possibles pour l'hybridation de l'apprentissage dans notre contexte.
- Choix des métriques en termes de coûts, de performances et de sécurité. Le coût de l'apprentissage doit être calculé en terme de temps de calcul et de communications. Il est également important de mesurer la qualité du résultat obtenu pour la détection. Et enfin, en termes de sécurité, les données utilisées sont des données sensibles : il est important d'identifier les risques, les attaques possibles et l'impact sur notre solution.
- Evaluation des différents modèles et comparaison avec le modèle centralisé : en fonction des métriques définis, nous pourrons évaluer les différents modèles d'hybridation proposés et les comparer avec une solution complètement centralisée.

References

[1] Yves M Galvão, Letícia Castro, Janderson Ferreira, Fernando Buarque de Lima Neto, Roberta Andrade de Araújo Fagundes, and Bruno JT Fernandes. Anomaly detection in smart houses for healthcare: Recent advances, and future perspectives. *SN Computer Science*, 5(1):136, 2024.

[2] Daniel Joseph Warrington, Elizabeth Jane Shortis, and Paula Jane Whittaker. Are wearable devices effective for preventing and detecting falls: an umbrella review (a review of systematic reviews). *BMC Public Health (Web)*, 21(1):1–12, 2021.

[3] Qin Ni, Ana Belen Garcia Hernando, and Iván Pau De la Cruz. The elderly's independent living in smart homes: A characterization of activities and sensing infrastructure survey to facilitate services development. *Sensors*, 15(5):11312–11362, 2015.

[4] Pireh Pirzada, Neil White, and Adriana Wilde. Sensors in smart homes for independent living of the elderly. In *2018 5th International Multi-Topic ICT Conference (IMTIC)*, pages 1–8. IEEE, 2018.