IOT-ML : Secure Machine Learning on IOT Traces for Daily Activity Discovery

(PhD proposal or Master 2 Internship – including part-time trainee – leading to a PhD on the same topic)

Supervision : Luc Bouganim (Luc.Bouganim@inria.fr), Philippe Pucheral (Philippe.Pucheral@inria.fr)
Applications via email (CV, motivation letter).

Context.
The PETRUS team (Inria/UVSQ), in association with the Hippocad company and the Yvelines Departmental Council, is currently deploying secure home boxes for 10,000 patients in the Yvelines region. These boxes, based on the team's research results (DBMS embedded in secure hardware), include a personal medical-social database to improve care coordination for dependent people at home. Medical and social workers interact with these secure boxes via a smartphone application. Our objective is to enhance these boxes with the ability to communicate with IOT sensors measuring e.g., luminosity, movement, temperature, to improve patient monitoring. The raw data from the sensors are analyzed by Machine Learning (ML) techniques to identify the patient activities and thus, detect the evolution of patients towards risk situations like depression or illness. Because of their precision, these raw data are however very intrusive. The originality of our approach is to allow a local processing of these data in each box which includes hardware security elements, in order to externalize only the relevant information: alerts, aggregated values on patient dashboards.

Challenges.
ML algorithms build a model based on a training dataset in order to make predictions, in our case, to discover the activity of an individual based on her IOT traces. Beside the classical issues of data representations (from IOT traces to a dataset that can feed an ML algorithm), our approach faces two challenges:
First, we have no possibility to obtain a training dataset for each targeted home-box user. Indeed, we cannot ask elderly people to label their activities during some weeks in order to build the corresponding training datasets: It would be too complex, costly and error prone without a personal assistant. We can however use existing datasets labelled for daily activity discovery (e.g., [1]) and use semi-supervised ML approaches [3] to dynamically adapt the produced model to the targeted home-box user. Indeed semi-supervised approaches use un-labelled data to refine an existing model obtained on labelled data. Other strategies could be defined based on a minimal feedback from the user or on some questionnaires describing the typical activities of the user.
Second, the ML algorithms must be computed inside the home-box, and more precisely in the secure part of the home-box which is composed by a microcontroller with limited RAM resource and a trusted platform module (TPM). Thus the algorithms must be efficient despite limited RAM resources. This may imply to define specific data structures adapted to this environment.

Methodology and expected results
- State of the art of relevant ML strategies, supervised learning, semi-supervised learning
- Analysis of a real dataset (labelled activities during one month) and on a synthetized dataset (un-labelled). Data will be transformed thanks to simple automata (already existing).
- Selection of an ML algorithm and adaptation to our context
- Implementation and test on the proposed datasets
- Proposition of a data model to store efficiently the IOT traces (in adequation with the chosen algorithm)
- Modification of the chosen algorithm to take into account the scarce resources

Applicant's expected background.
- ML algorithm knowledge
- Python (knowledge in C or Rust will be appreciated)
- The applicant could be willing to do a Master2 internship or a part-time trainee (Master2 level), or having completed a Master2 and willing to do a PhD

References.
[1] Alemdar, Hande & Ertan, Halil & Incel, Ozlem & Ersoy, Cem. (2013). ARAS Human Activity Datasets in Multiple Homes with Multiple Residents. 232-235. pdf
[2] Jérémie Saives, Clément Pianon, Gregory Faraut. Activity Discovery and Detection of Behavioural Deviations of an Inhabitant from Binary Sensors. IEEE Transactions on Automation Science and Engineering, Institute of Electrical and Electronics Engineers, 2015, 12 (4), pp.1211 - 1224. pdf
[3] van Engelen, J.E., Hoos, H.H. A survey on semi-supervised learning. Mach Learn 109, 373–440 (2020). pdf