Engineer / Postdoc Proposal

A Framework for Sensor Data Management and Analysis

Current team: K. Zeitouni (Professor, UVSQ/DAVID Lab), Y. Taher (Ass. Professor, UVSQ/DAVID Lab), Z. Kedad (Ass. Professor, UVSQ/DAVID Lab, HDR), C. Gouy-Pailler (Researcher, CEA-LIST)

ABSTRACT

Sensors and actuators play a key role in the development of many sectors such as smart cities, industry 4.0, environment and health. They provide an unprecedented opportunity to monitor, study, and interact with the physical world.

Modern sensors are always connected and are able to continuously report various parameters. With a large deployment of sensors, the amount of data generated and their velocity exceed the capacity of traditional data management and analysis systems.

The objective of the SDMA project is to design and implement a framework specifically tailored for sensor data management and analysis. It aims at covering a process workflow, ranging from stream data management, cleaning, querying and aggregation, enrichment with contextual and semantic data sources, high-level event modeling and prediction, and decision support. Finally, we will provide visualization tools like dashboard to allow monitoring and synthetic analysis of the observations in real-time.

To achieve this objective, we will leverage data mining and machine learning algorithms in the context of sensor readings. Furthermore, we intend to combine sensor data management with complex event processing techniques in order to trigger actions or recommendations, or to raise alerts.

We target a generic or easily customizable framework for various applications dealing with the Internet of Things and smart cities. Our use case lies in the context of mobile participatory sensing in relation with the ANR project POLLUSCOPE (www.david.uvsq.fr/polluscope), which uses low-cost wearable sensors hold by volunteers along their daily activity to collect and analyze their exposure to air pollution.

This postdoc proposal aims at participating in the design and the development of this framework. In particular, the main missions will be: (i) the refinement of the architecture design using up-to-date concepts and technologies (namely micro-services and containers); (ii) the study of multivariate time series (MVTS) analysis in the context of IoT, in order to select adequate data mining/machine learning algorithms for high-level knowledge extraction; (iii) prototyping and evaluating the framework, showing its effectiveness (based on the scenarios of Polluscope) and its genericity (by using public datasets, e.g., “wafer” dataset to showcase a smart manufacturing scenario).

The development of MVTS analysis should form part of the open source initiative (e.g., scikit-learn) supported by the CDS project.

Duration: 12 month

This project involves multi-disciplinary collaborations with scientists in environment (AIRPARIF, UVSQ/LSCE), and health (INSERM).

Potential collaboration with the IRS ACE-ICSEN.