Fondation Mathématique Jacques Hadamard

Université Paris Saclay

Research initiative in industrial data science (IRSDI)

Call for projects – March 2019

Overview and scientific scope

The Research initiative in industrial data science (IRSDI - *Initiative de Recherche en Sciences des Données pour l'Industrie*) is a corporate patronage funded by EDF and THALES and operated by the Jacques Hadamard Mathematical Foundation (FMJH - *Fondation Mathématique Jacques Hadamard*).

It is part of the Gaspard Monge Program for Optimization, operational research and their interactions with data science (PGMO - *Programme Gaspard Monge pour l'Optimisation, la recherche opérationnelle et leurs interactions avec la science des données*), launched by EDF and the FMJH, and with Thales and Orange as current additional partners.

The focus of this IRSDI initiative is on data science and the many ways it may help industry, within the fields of energy and complex systems for surveillance and defense domain.

Mathematicians and computer scientists from both the academic and industrial worlds can benefit from it. Projects are open to all academic researchers with no restrictions due to administrative or geographic location. Nevertheless, small teams with few but committed researchers will be favored.

Projects to be funded should be relevant to the field of data science (including machine learning, statistics, optimization and computer science in relation to data analytics) and should be

focused on solving industrial problems in the fields of energy or complex systems. A list of suggested topics is included in appendix.

Objectives

The objective is to support research projects through collaborative actions between academic researchers and industrial researchers or practitioners, focused on solving industrial problems in the fields of energy or complex systems. These projects are encouraged to be a kick-off for a future partnership between academic and industrial researchers.

Each proposal is thus formed by a pair given by an academic team and a partner company. The academic team must clearly identify a scientific leader, whose lab will manage the funding for the rest of the team. The industrial partner is not eligible to funding.

The partner company must identify a corresponding member and will have to write a support letter describing the industrial challenges to be addressed, the data sets to be studied and the expected benefits of the collaborative research to be undertaken. *The true research (and not only development) nature of the project should be underlined in this letter.*

The partner companies do not necessarily need to be EDF or THALES, though these two companies are extremely willing to build partnerships through this sub-program.

Call for projects: schedule

- March 2019: publication of this call for projects
- Before final submission: prospective candidates are encouraged to get in touch with the PGMO board (via Gilles Stoltz, gilles.stoltz@math.u-psud.fr) to get some pre-submission feedback on the proposal
- May 13, 2019: deadline for submission of the projects (link indicated below)
- July 2019: notifications of acceptance or rejection to the project leaders (after recommendations issued by the scientific committee and final decisions made by the executive committee of the PGMO)

Submission of projects: via EasyChair, at the URL
https://easychair.org/conferences/?conf=pgmo2019

Template: a submission template is provided at
https://www.fondation-hadamard.fr/fr/pgmo-calls-projects/call-project
(Note that a single PDF file describing the scientific content of the project as well as all required administrative information is expected; it can be written in French or in English.)

Call for projects: rules

What follows is only a summary of the general PGMO submission rules, fully detailed at
https://www.fondation-hadamard.fr/fr/pgmo-calls-projects/call-project

Necessity of an industrial partner; note on data and on the codes

An IRSDI project consists of a pair composed by an academic team and an industrial partner.

Projects must emphasize the link with real data. Projects based on public/open data or on the creation of public/open data resembling to industrial or confidential data will be particularly welcome. If the data sets to be studied need to be collected and created first, the project leaders must describe the methodology to be followed and provide a timeline.
Codes are encouraged to be made available publicly.

Funding expectations / Budget rules

Projects should typically last 1 year (typically, October 2019 – November 2020).
Funding per project will be typically from 10 to 15 kEuros, but might reach, in specific cases, 20 kEuros. We expect to fund about 6 or 7 projects.

All typical research expenses such as travels, computers, internships (or other non-permanent staff), invitations of researchers, purchase of data, etc., can be covered. Upon funding, an agreement will be signed between the main lab for the project and FMJH, and the lab will handle the obtained money. This lab or research institution or teaching institution must be from the academic world.

Commitment by funded project teams

All funded projects will be asked to participate, at the end of the project, to the annual review of IRSDI projects within the PGMO workshop in Fall 2020 (typically lasts one morning and part of the afternoon).

Support by PGMO / IRSDI will have to be acknowledged in publications relative to funded projects.

A follow-up committee composed of representatives of the funding companies EDF and Thales will get in touch with and may visit the project teams during the 2019-20 year.

Contacts

From the PGMO executive board:

- Gilles Stoltz (CNRS / Université Paris-Sud, gilles.stoltz@math.u-psud.fr)

PGMO / IRSDI industrial sponsors

(May help to build projects by pointing to interested members of EDF and Thales)

- Georges Hébrail (EDF, georges.hebrail@edf.fr)
- Anthony Larue (Thales, anthony.larue@fr.thalesgroup.com)

Scientific committee of PGMO: Its composition can be found at the bottom of the page https://www.fondation-hadamard.fr/fr/pgmo

List of suggested topics

Several topics are welcome, including but not limited to the following ones (listed with no order of priority). They were suggested by the industrial sponsors EDF and THALES, in the fields of electricity production, optronic systems, and C2 (Command and Control) Centers.

Advanced processing and classification for surveillance system (optronic or radar)

- Detection and tracking of objects in image or radar
- Times series (potentially multidimensional) and images (potentially hyperspectral) classification or prediction with supervised or unsupervised approaches
- Simulation/synthesis of images, times series or data using machine learning algorithms (GAN, style transfer…)
- Efficient strategy of data labelling to build large dataset
- Tools for automatic parameters processing chain settings using performances criteria
- Incremental and few shot learning
- Indexation and search engines or association engine for images or times series: Context Based Images retrieval, similarity search
- Times series or graphs database and indexation

Optimization for system design and management

- Processing algorithm and optimization for cooperative system (system of systems): optimization of radar network or radar waveform…
- Predictive maintenance and fleet management using statistical model, machine learning, expert system…

Analysis and forecasting of energy related time series

- Visualization of energy consumption time series (households, buildings, electrical networks)
- Electricity load and production forecasting at small scale and short term; probabilistic forecasts; forecasts using large scale exogeneous data; spatial forecasts; new

optimisation methods for energy forecasting (reinforcement learning, modern optimization methods for variable selection, multi-task learning)

- Forecasting of renewable production and real-time self-consumption optimization: optimal paving of an area with load patterns under constraints
- Disaggregation of electricity consumption by usage at individual level or small aggregates (e.g., household/building, NILM – Non-Intrusive Load Monitoring)
- Image-based load profile analysis of electrical appliances to detect and identify failure / degradation over time / different work modes
- Privacy-preserving analysis of individual energy consumption (clustering, forecasting, scoring on anonymized/encrypted data)
- Generation of synthetic data featuring electrical networks (network topology and load in operation)
- Real-time data analysis and online learning for optimal decision in energy markets
- Multi-scale spatio-temporal analysis of multivariate data, analysis of time series connected with unstructured data (e.g. text)
- Multi-source data assimilation: physical models, high quality sensors data and crowd sourced data sets

Maintenance of power plants, electrical networks and consumption premises

- Integration of various data sources (description of equipment, context of operation, maintenance operations, monitoring information, alarms, logs, sensor values, computation results from physical models and simulation codes, expert knowledge, etc.) based on semantic web approaches like RDF/OWL;
- Visualization of multi-source, heterogeneous and multi-sensor data
- Monitoring, diagnostics and prognostics approaches using predictive models and based on the exploitation of complex and heterogeneous data (graph data, textual data, multi-dimensional time series from sensors, images, video, physical models, expert knowledge, etc.).
- Analysis of signals or images coming from non-destructive testing of complex material (concrete, complex geometry or complex degradation type);
- Image recognition and indexing for power plants inspection, building energy efficiency analysis, photovoltaic power generation forecasting, detection of tree pruning needs around electrical networks

- Forecast and identification of crises in electrical networks from past data, models to predict impact of weather events
- Hybrid models (physical, statistical, machine learning) to estimate electrical component reliability in electrical networks
- Certification, verification and validation of machine learning methods in predictive maintenance
- Hybrid approaches combining physical models, learned models and expert knowledge

Exploitation of textual data in the energy domain

- Text generation in Customer Relationship Management (CRM): automated response suggestion for email and chabot, description of consumption time series
- Automatic text summarization: call-center conversations, power plants maintenance reports

Optimization and control of complex systems

- Demand-response management (optimal sequential selection of customers who are asked to reduce their consumption in order to facilitate load balancing)
- Multi-energy adaptive control of a territory using machine learning approaches
- Tight coupling of stochastic optimization and stochastic forecasts of electricity load, application to unit commitment, dispatch, capacity expansion planning
- Markov Decision Process to learn optimal control of unknown dynamics, or strategies for repeated complex or combinatorial optimization (power plants involving uncertainty such as weather or changing performance of machines, smart home control, optimal core design of nuclear power plants, energy planning smart grid management involving multi energy systems and electric vehicles).