1SC4195 – Data-driven reliability estimation and optimal operation planning for health care equipment

Instructors: Sorin Olaru
Department: DÉPARTEMENT GÉNIE INDUSTRIEL ET OPÉRATIONS
Language of instruction: ANGLAIS
Campus: CAMPUS DE PARIS - SACLAY
Workload (HEE): 40
On-site hours (HPE): 27,00

Description

Health care equipment in general has high reliability requirements: its failure might directly endanger the lives of patients. At the same time, health care equipment also has high availability requirements: the hospitals in general could not afford too-long downtime as they have to keep serving patients. Understanding the reliability of health care equipment is, thus, an important topic in medical care industries.

In this project, we work with GE Healthcare (GE HC), one of the leading suppliers to health care equipment globally. GE HC has to satisfy the high reliability and availability requirements for their products. For this, they have to maintain a large-scale aftersales(service) supply chain supporting over one million systems installed globally in the world. More than 400,000 spare part references, including~10,000 repairable parts are potentially necessary to maintain the installed based (IB). How to effectively manage so many products, while satisfying the high reliability and availability requirements, is, thus, a very challenging problem. At the same time, during the product design, development, and, more importantly, the operation phases, there are a large number of data available. Although these data might be noisy and contain large degree of missing information and uncertainty, they are valuable sources that could provide some insights to the reliability of the products, which could be further used to improve the efficiency of the operation of the after-sale supply chain.

Quarter number
ST4

Prerequisites (in terms of CS courses)
Statistics and probability

Syllabus
The project will apply tools and processes to leverage the reliability information from the data and to enrich decision making process in the

operations of the after-sale supply chain. In this project, the students will be able to:
- Work with the real dataset provided by GE HC;
- Develop data-driven reliability models for health care equipment, on both the component and system level;
- Experience how to deal with the “imperfectness” of a practical dataset;
- Experience how to improve a current process with the help of data analytics on the reliability data.

Class components (lecture, labs, etc.)
Project-based learning

Modalities of participation of the partner involved
o Participation of speakers GE within the introductory session.
o Presence of an expert from GE in coaching mode during the week.

For the project, industrial data (possibly anonymized) will be provided to students. Some will be labelled, others not. Some will include historical information from the maintainers: types of defects or degradations identified, maintenance actions carried out), others will not.
Students will be asked to combine knowledge of physical systems and data processing algorithms to extract useful information from the raw data and make recommendations for use by maintainers.

Two modalities are envisaged:
- Either divide the tasks into small groups of students;
- Either give the same tasks to two or three groups that will work in parallel, in challenge mode.

Grading
Oral presentation and evaluation of the project along the week

Course support, bibliography
References of the ST4 GSI lectures

Resources
1 room for 40 students, overhead project, partition by groups
Software : Matlab/Simulink (network licence – wifi preferably) on individual PCs for each student; Python, …
2 academic supervisiors + 1 expert from GE
Involvement of GE over the entire week when needed (upon request)

Learning outcomes covered on the course
At the end of this course the students will be able to:
- to understand the issue of resilience and predictive maintenance as well as the potential benefits of such an approach
- to analyse a complex system and develop the elements of reflection

leading to a representation model (physical, statistical, etc.) of the phenomena whose identification is the subject of the study
- to conduct the modelling process with an appropriate choice of modelling assumptions and to understand the limits of the models
- apprehend and use machine learning algorithms to extract information useful for predictive maintenance from raw data
- conclude and decide on the relevance of the approach and on the performance of the algorithms evaluated

Description of the skills acquired at the end of the course
C1: Analyze, design and build complex systems with scientific, technological, human and economic components
C.1.1: The aim is to understand and analyse a complex system and study the problem as a whole for decision-making where the economic and human dimensions are important.
The cost and human impact aspects are important for decision-makers, who will have to convince them of the economic benefits of such a solution while maintaining a high level of security or even improving it depending on the project, and to convince users of the viability and reliability of the solutions envisaged.
C1.2: Use and develop appropriate models, choose the right modeling scaleand relevant simplifying assumptions to address the problem
In all projects it will be necessary to recover a large volume of data from an industrial company and from different sources. This will raise the problem of data aggregation, assumptions for eliminating outliers and model(s) for selecting the appropriate scale and level of accuracy required.
C4: Ability to create value for your company and its customers
In this context, it will be a question of creating value for the customer by increasing availability time by assessing the risk of failure and optimizing maintenance phases (periodicity, critical equipment, etc.).
C6: Being comfortable and innovative in the digital world
C6.5 Use any type of data, structured or unstructured, including massive data.
This will involve processing massive data provided by the manufacturer. This data, exploited using algorithms to be developed, will provide information on the risks of failure. The use of data for diagnosis will require prior treatment to eliminate outliers.
C7: Collect relevant and reliable information to support an argument (e. g. to retrieve relevant data).
C8 Lead a project, a team

ST4 – 42 – BIG DATA & HEALTH: FROM DATA ACQUISITION TO DECISION MAKING

Dominante : VSE (Living-Health Environment)
Langue d'enseignement : French
Campus où le cours est proposé : Paris-Saclay

Engineering problem

Health, and more generally the field of living organisms, is undergoing a veritable technological revolution that is making it possible to acquire increasingly large amounts of data (signals, images, measurement results, etc.). For example, genomic data can now be used to explore the multi-scale activity of a cell in ever greater detail. Another example is medical imaging, which is one of the pillars of medicine, whether in conventional medicine and surgery or in the neurosciences. The "data revolution" is impacting all areas of life and we propose, within the framework of this thematic sequence, to explore the consequences in epidemiology, oncology and in the context of neurodegenerative diseases. This thematic sequence is resolutely positioned on the processing of biomedical data and offers an area for meetings and exchanges between doctors/biologists from various backgrounds and engineers. This thematic sequence offers a field of application where statistics and machine learning find all their meaning.

Adviced prerequisites

None

Context and issue modules: This introduction to the thematic sequence will discuss the place of data in the health field and highlight the engineer's position in this context. Data collection, whether for large cohort studies or clinical trials, requires innovative equipment: on-site visits to one of our partners will be offered to enable you to understand this important aspect prior to any statistical study. The different integration lessons will also be presented by the partners.

Specific course (60 HEE) : Biology and Statistics

Brief description: The cell is a complex multi-scale system that biologists have been studying for several centuries. The development of new technologies now allows us to explore it and measure its activity more and more finely. The objective of the first part of the course is to present the

various facets of the cell to better understand its global functioning. The technologies of measurement will also be approached.

The second part of the course presents the statistical methods widely used in the biomedical field. At the end of this part, the students will have a useful toolbox to answer the various questions raised by the analysis of biomedical data.

This course will be largely illustrated on medical applications in order to allow students to acquire the knowledge and vocabulary necessary to understand the applications proposed in IE (bases in biology, genomics, etc.), to understand the origin of the enormous volume of data related to medical applications. For this reason, about half of the overall volume of this course will be reserved for our physician/biologist partners.

Challenge week n°1: Sleep analysis by electroencephalogram
- Associated partner: DREEM & Institut du Cerveau et de la moelle épinière (ICM)
- Location: Paris-Saclay campus
- Brief description. In the framework of this integration course, we propose to implement signal processing and machine learning methods to address neurological disease issues. In particular, we will be interested in the study of sleep disorders through the analysis of electroencephalogram (EEG). These disorders are generally early signs of neurological disease (such as Parkinson's disease), hence the importance of monitoring them. This is therefore a problem of EEG signal processing and machine learning where we will try to produce reliable biological/medical knowledge from massive noisy and incomplete high dimensional data.
Course of action: The students involved in this project will be divided into sub-groups which will compete in a "challenge mode": The sub-groups will analyse the data through the different techniques and compete to produce the best model. The models will be evaluated in real time and a live ranking will be provided and visible to all. A collaborative platform will be used to share scores, results and scripts. As the data and the problem are provided by our external partners from DREEM and ICM, the students will have to meet them several times. Finally, they will have to present their results in front of all the actors of the project (all the students involved in the project, partners, supervisors).

Challenge week n°2: E3/E4N cohort data for the identification of major trends
- **Associated partner:** INSERM/Gustave Roussy
- **Location:** Paris-Saclay campus
- **Short description:** In this integration course, we propose to implement statistical/machine learning methods to answer epidemiological problems. In particular, we will be interested in the data of the E3N/E4N cohort (Epidemiological Study of MGEN women), which concerns

hundreds of thousands of women (https://www.e4n.fr). This is therefore a problem of statistical analysis of a "Big Data" dataset where we are trying to identify major trends at the population level from massive noisy and incomplete data.

- Process. The students involved in this project will be divided into sub-groups. The subgroups will analyse the data using the techniques of their choice. The data and the problem being provided by members of INSERM, the students will have to meet them several times. Finally, they will have to present their results to all the actors of the project (all the students involved in the project, partners, supervisors).

Challenge week n°3: Infectious diseases: The indoor environment project

- **Associated partner:** Institut Pasteur
- **Location:** Paris-Saclay campus

Brief description: The "Milieu Intérieur" project, coordinated by the Institut Pasteur, has been developed to better understand the diversity of immune reactions in the population with the broader aim of contributing to the development of precision medicine. In the framework of this project, a cohort of 1000 individuals was constituted and several large datasets were generated, including socio-demographic and clinical data, blood cell composition data, as well as gene expression data under different immune stimulation conditions for each individual. During this IR, we propose to analyse and cross-reference these data in order to identify factors that differentiate individuals in their responses to immune stimuli and eventually to explain and predict these responses. The objective will be to apply descriptive and predictive statistical analysis methods to extract relevant biological information from these data.

Procedure. The students involved in this integration course will be divided into subgroups. The subgroups will analyse the data using the different techniques discussed in class. As the data and the problem are provided by the Pateur Institute, the students will have to meet this partner several times. Finally, they will have to present their results to all the actors of the project (all the students involved in the project, the partner, the supervisors).

Challenge week 5: Using innovative methods to predict health and disease in exposome studies

Associated partner: one of the institutes involved in the European ATLHETE project (https://athleteproject.eu)

- **Location**: Paris-Saclay campus
- **Brief description.** The exposome, described as "the totality of human environmental exposures from conception onwards", recognises that individuals are exposed simultaneously to a multitude of different environmental factors and takes a holistic approach to the discovery of etiological factors of disease. The main advantage of the exposome over traditional "one exposure-one disease" approaches is that it provides an unprecedented conceptual framework for the study of multiple

environmental risks (urban, chemical, lifestyle, social) and their combined effects. The objective of this AR is to promote statistical, data science or other innovative quantitative approaches to study the effects of exposure indicators (exposome) on health. The dataset will include multiple health phenotypes (quantitative or qualitative), multiple exposures, -omics and additional non-exposure variables (e.g. potential confounders).

- Process. The students involved in this project will be divided into subgroups. As the data and the problem are provided by our partners in the European ATHLETE project (see https://athleteproject.eu/consortium), one of these partners will participate in the supervision of the project. The students will have to meet him several times. Finally, they will have to present their results in front of all the actors of the project (all the students involved in the project, partners, supervisors).

Challenge week 6 : Epidemiology & Biostatistics on a Hospital Health Data Warehouse

- **Associated partner:** AP-HP (Assistance Publique - Hôpitaux de Paris)
- **Location:** Paris-Saclay campus
- **Brief description:** The Health Data Warehouse (EDS) of the Assistance Publique - Hôpitaux de Paris (AP-HP) integrates the administrative and medical data of more than 8 million patients hospitalised or consulted in one of its 39 institutions. This warehouse allows for the improvement of hospital activity management and the advancement of scientific research in the field of health by promoting the implementation of studies on real-life data, the setting up of clinical trials and the development of decision support algorithms. In this Integration Course, we propose that you discover the AP-HP's EDS and its specificities through the prism of epidemiological and biostatistical analysis. In particular, you will have the opportunity to process structured data (medico-economic data - diagnoses & procedures, demographic data, patient pathways and biological analyses) and unstructured data (medical reports), integrated into a synthetic database largely inspired by the structure of the EDS. You will have to answer two research questions (epidemiological surveillance study and analysis of predictive factors for lung cancer), while deciphering and correcting the biases inherent in routine clinical data.

Process. The students involved in this project will be divided into subgroups. The subgroups will analyse the data using the different techniques discussed in class. As the data and the problem are provided by the AP-HP, the students will have to meet several times with this partner. Finally, they will have to present their results in front of all the actors of the project (all the students involved in the project, partner, supervisors).