Award Ceremony at Congress Graz

Prize money
€ 6.000,-

1st LEC DATA CHALLENGE

> Hunt down the fault

1.5. - 19.8.2019

Background

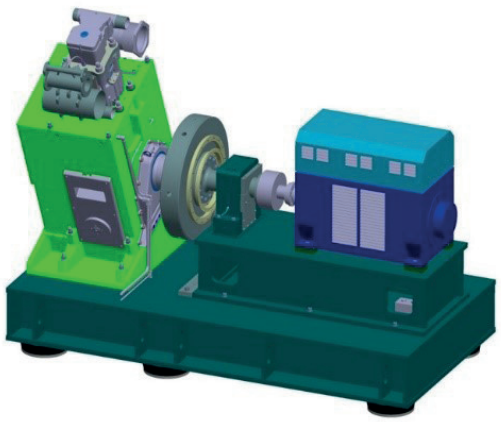
The LEC focuses on developing environmentally sound, efficient and robust large engines for sustainable transportation and energy solutions. In the research and development process, systematic measurements (e.g. on research engines, see figure) are indispensable for evaluating technological advances. To conduct measurement campaigns as efficiently as possible, it is necessary to check the quality of the data immediately after it is generated directly at the engine test bed. In the event of a measurement fault, the prompt detection of the fault helps to ensure that corrective action is taken – e.g., the replacement of a faulty sensor.

Methodology

There are multiple ways to detect measurement faults, from simple limit checks to model-based methods that consider the physical relationships between the variables. Since physical modeling of highly complex systems is limited, however, fault detection methods are required that are based on historical data instead of explicitly physical knowledge. With our first LEC Data Challenge, we would like to encourage data-enthusiastic students and practitioners from a variety disciplines to develop data-based fault detection methods that are best suited for our large engine applications.

Data

Our challenge provides two datasets, which are included in the file challenge.zip. One dataset is the training dataset (training_dataset.csv). Along with fault free data, it contains two groups of faulty measurement records (i.e., a series of consecutively recorded measurement points that is subject to measurement faults). To illustrate which measurement records and measurement channels are faulty, the file training_solution.csv is provided. The values in this file are 1 or 0, with 1 corresponding to „faulty“ and 0 corresponding to „fault free“. The other dataset is the test set (test_dataset.csv) that contains a certain number of faulty measurement records which data challenge participants should identify. In this test set, the spread of faulty measurement records (overall number of faulty measurement records and grouping of consecutively recorded measurement points subject to the same fault) as well as the intensity of the faults (fault amplitude) are different from those of the training dataset. The two datasets include a total of 639 measurement records of 15 variables.



Challenge

Determine which data points of the test dataset are faulty and which are fault-free (channel and record number).

Submission

Results must be submitted as a comma separated value (CSV) file with exactly 300 rows and 15 columns. The values must be 1 or 0, with 1 corresponding to “faulty“ and 0 corresponding to „fault free“, similar to the training_solution.csv file of the training dataset. Please submit your results (test_solution.csv) via e-mail (data.challenge@lec.tugraz.at). As an incentive, it is possible to make a voluntary intermediate submission and obtain feedback on how close you are to the goal. The deadline for the intermediate submission is July 15, 2019; the fi nal deadline is August 19, 2019.

Evaluation

The intermediate results and the final results will be ranked based on the number of correctly classified data points (faulty or fault-free). Thus, it is important not to identify fault-free records as faulty or vice versa. The best possible performance is indicated by 100 percent correctly classified points. The evaluation results will be sent to the email address you have provided. The feedback on the intermediate results will be sent by July 29, 2019, and the final evaluation results by September 5, 2019. The top five teams will be invited to present their contributions at the 17th symposium “The Working Process of the Internal Combustion Engine”, which focuses on sustainable Mobility, Transport and Power Generation and takes place in Graz on September 26-27, 2019. The three best solutions selected by a jury of experts will receive attractive prize money.

What is also important

Participation in the challenge is possible from May 1 to August 19, 2019. Updates will be announced at the website www.lec.at/datachallenge.
If you have any questions, please contact Doris Schadler at data.challenge@lec.tugraz.at