SBB Fibre Optic Sensing, Train detection models

Degree programme: MAS Information Technology

Over the years, improvement the monitoring capabilities over civil infrastructures has became a priority to avoid services interruptions and reduce the maintenance cost. In this context the SBB started a pilot project to evaluate the capabilities of FOS technology. This master thesis addresses the goal to detect the presence of a train in the FOS signal, as starting point to identify different incidents over the railway, like; rock falling, railway breaks, cable theft...

Context

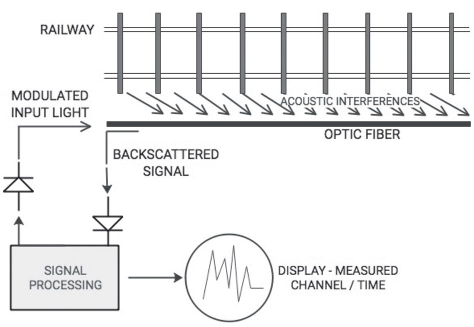
The Fibre Optic Sensing technology uses commercial fibre optic cables, as a distributed sensor. Using the backscattering property of the light and Coherent Optical Time Domain Reflectometry, a special hardware unit sends a pulse of light into a fibre optic, this pulse is naturally scattered due to the imperfections of the fibre and it is also affected by tiny strain events, which are determined by localised acoustic environments. Reading and digitalizing this signal, a large infrastructure like a railway, can be monitored.

Data description

The signal is stored continuously for monitoring. The SBB delivered a set of 7 H5 files of 2 Gb to address an analysis with the goal of detecting the position of the train. Each file contained 5 minutes of measurement in a matrix format, with 2724 rows and 750000 columns, in which the rows represent discrete positions along the cable and the columns represent the measures on each position through the time.

Data preparation

Following the CRISP-DM methodology, an analysis of the files, using Jupyter-Notebooks was conducted. After a descriptive analysis I realized that the noise of the train produces high frequency interferences on the scattered light signal. By using a high pass filter, the signal interference produced by the train could be delimited an over a heat map visually represented.



FOS Technology schema for Railway monitoring.

Models

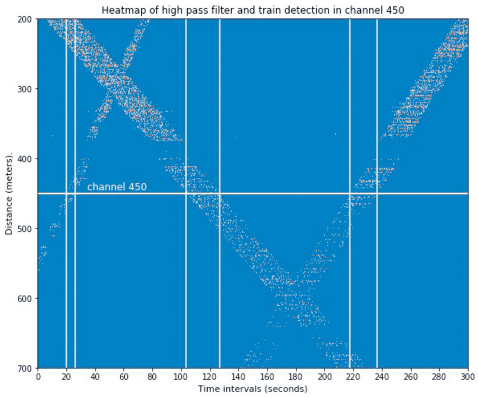
Three different models were developed to determine the existence of a train in a window of 1 second, one based on digital signal processing and two based on learning machine algorithms.

- **PEAK Model**: Based on counting of local maximum points of the signal above a threshold.
- **Random Forest Model**: The Random Forest is an ensemble of 10 decision trees classifiers that after a voting process classifies the window in two classes 0=No Train, 1=Train.
- **Support Vector Machine**: It is another type of classifier trained with the stochastic Descent Gradient algorithm.

Once the models were built, they were included in a process to delimited the train areas given a complete channel signal, previously filtered. To evaluate and compare the three models their confusion matrixes were build.

Conclusion

The lowest error rate was achieved with the Random Forest model; it demonstrates its possible usage to identify random patterns and with a correct training set be able to address the identification of different railway incidents.



Heat map representation of the train traces and the limits the train area on channel 450.

Diego Martín Barreiro Fandiño
+41 78 945 03 88
dmartin.barreiro@gmail.com