Pipeline leak detection using methods of artificial intelligence

Author: Mgr. Lukáš Grofčík
Supervisor: doc. Mgr. Martin Drlík, PhD.

Constantine the Philosopher University in Nitra
Faculty of Natural Sciences and Informatics

1. Motivation

In recent years, data mining and predictive analytics are becoming a growing trend amongst industrial enterprises. In order to ensure the quality of their products and services, companies use machine learning to predict potential problems or risks.

Using analysis of data gathered from pipeline sensors alongside an artificial intelligence model may enhance the accuracy and speed at which companies can discover occurring leaks. Such a solution could help **reduce the risks and potential costs** of a company and ensure the **safety of environment** as it could prevent the leak of a fluid that can be environmentally dangerous or costly in resources.

The goal of our work was to examine the possibility of catching leaks occurring in pipelines with fluids using neural network Autoencoder and afterwards to create an automated process capable of handling all processes involved.

3. Automatization

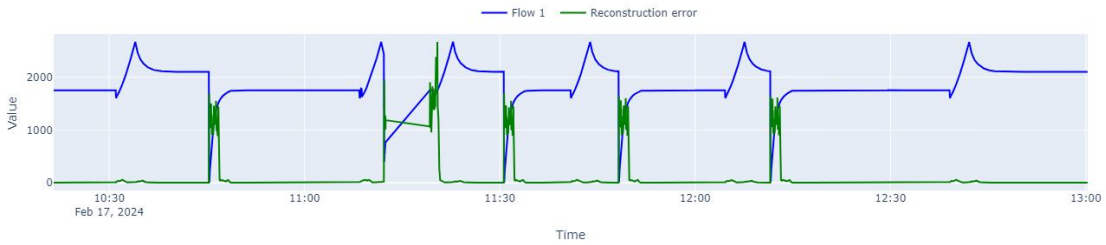
In order to be able to create the automatization of processes we firstly built an application capable of using our trained Autoencoder model. Such an application consisted of a **Python application** which handled the usage of model for predictions and a **PHP application** which handled the interaction with database, Python server and receiving values of sensors from pipeline.

The automatization itself was built using **GitHub Actions**. By creating two separate workflows, each using 4 jobs. We were able to create 2 **automatization processes** for initial training of model and retraining of model in case it starts to drift.

2. Modeling Autoencoder

Using a pipeline simulator we have generated a set of raw data containing values of four sensors (two for **flow**, two for **pressure**) with two being at the beginning of a pipeline and two at the end. After cleaning the raw data, we have aggregated them by time into rows of 400 values (100 for each sensor). This way our prepared dataset comprised rows of **time series** each containing sequential data for **100 seconds**.

Using this dataset we trained an **Autoencoder** model to learn the **patterns** of pipeline sensors data during time when no leaks are present. Afterwards we have tested and evaluated the **reconstruction error** of said model to assess the ability to correctly evaluate whether or not a leak is happening.



4. Results

An autoencoder model created in our work was capable of assessing leaks in cases where the pipeline pump was not used. In cases where the pump was used, the model was not able to differentiate between leaks and the usage of pump.

UNIVERZITA KONŠTANTÍNA FILOZOFA V NITRE
PRÍRODNÝCH VIED A INFORMATIKY