

## **Labs 7-8: Project 1 – Prediction using NN and Fuzzy Systems**

(Week 4)

### **1 – Objectives**

In this work you will try to predict when will an IoT gateway device fail based on the requests it receives and its processor load. Most of the requests received by the device are simply related with routing, but some require data processing, and can be quite computationally demanding.

### **2 – Dataset**

We will use the pre-processed artificial dataset “IoTGatewayCrash”.

The IoTGatewayCrash.csv file is composed of 2000 records. Each record corresponds to a 20 minute period of operation of the device that contains two inputs and one output:

- a) Input 1: The normalized number of requests received by the device during the 20 min period.
- b) Input 2: The processor average load during that period.
- c) Output: Normal Operation (0) or Crash (1).

The dataset is highly imbalanced, and there are very few crash instances.

### **3 – Will it crash or not?**

This is not a simple classification problem since we are dealing with sequential data, and probably what happened before is an important factor on the failure of the device. Therefore, you will likely need to consider the past. Since there are not many crash instances, it is unlikely that a fully automatic unsupervised method will be able to learn anything useful out of the data, but why not try?

Start by implementing a MLP with 2 inputs and 1 output and see if you can obtain any interesting result. Decide which hyperparameters to use (but don't waste too much time optimizing them), and evaluate the performance using Precision, Recall and F-measure.

You will likely reach the conclusion that you will probably need to use more than this kind of brute force approach. So, look at the data and see if you can get some insights on the problem, and what kind of features you should try to use as the inputs to your model...

### 4 – Enter the expert...

An expert in intelligent systems was contacted to give us some insights regarding this problem. Here is what he has to say:

[Expert1]

“A quick look at the data shows that this is a very imbalanced problem, with very few positive instances. So, you should try to balance your training set before any attempt to learn from the data” (Note: DON’T balance the validation or test sets, only the training set).

“I also performed some quick basic data analysis and found out that all crashes happen during a High processor load, even though this is not a sufficient condition for the crash to happen. However, it seems that the number of requests itself isn’t related with the crash. Therefore, what causes the crash must be related with something that happened in the past. This is definitely a system where order is important.” (Note: Maintain the order of data in the training, validation and data sets)

“So, my advice would be to try to learn something using Processor\_Load as one of the inputs, and use several previous instances of Number\_of\_Requests as additional inputs” (Example: try using the following features as the inputs for the NN: Processor\_Load; Number\_of\_Requests<sub>t-1</sub>; Number\_of\_Requests<sub>t-2</sub>)

Create models using the features suggested by the expert and properly evaluate and validate their performance (use the knowledge previously acquired in this course on how to deal with the data and the experimental setup in order to have a proper evaluation and validation).

### 5 – Let’s get another expert...

An expert in how these Gateways work, was contacted to give us some insights regarding this problem. Here is what he has to say:

[Expert2]

“The problem with these Gateways, is that they were not originally developed to perform heavy computations for long periods of time. However, using them exclusively to process simple routing requests is a waste of processing power. So it makes sense to take advantage of them in some other ways, and it’s natural to assign them for some edge processing tasks. However, when some jobs are carelessly assigned to the Gateway, the system can overload and it might result in subsequent crashes. Usually this is caused by jobs that hog some of the system’s resources for too long. When that happens, the Gateway usually starts refusing regular requests for no apparent reason. Such requests are usually repeated until they are accepted. After a while, the processor starts overworking, heat builds up, and the system crashes.

In order to detect when this is about to happen, I would try to look for an abnormal number of requests in the periods before the processor load gets noticeably high.”

Use this new information to create a model that performs better (instead of using the number of requests as the inputs for the system, try to create new feature(s) with them).

Properly evaluate and validate the model performance (use the knowledge previously acquired in this course on how to deal with the data and the experimental setup in order to have a proper evaluation and validation).

### 7 – Fuzzy Rule Based Expert System

Implement a Fuzzy Rule Based Inference System to predict the crashes based on the experts' knowledge. Take into special consideration the knowledge provided by Expert 2.

Evaluate the performance of the system when compared to the NN based approaches.

### 8 – Generalization

While you have been developing the models, new data has been collected, so let's see how well each of the developed approaches perform on this new data. Use the new dataset as the Test Set. Do not retrain the systems! Simply test and evaluate them in order to see if they are able to generalize what they learnt.

### 9 – Report

Write a succinct report where you indicate the options you made regarding the data preparation, the experimental setup, the construction of each model, and all the evaluation and validation processes during all the phases of the project.

Note that I will be more interested in your reasoning and your options than in the end results.

Upload the report and a .zip containing all developed code via Fenix until Friday, **October22<sup>nd</sup>**, at 23:59.

The project will be discussed during week 8, and you will need to demonstrate your results with another test set that will be made available during the discussions (in order to properly test generalization).