MICROSOFT AZURE MACHINE LEARNING FOR DATA BASED MODELING

kateryna.mishchenko@se.abb.com, Corporate Research Center, ABB

September 26, 2017

Introduction

Process model development and optimization is one of the prioritized directions of research for the companies active in automation industry. One particular application of interest of ABB Corporate Research Center is automation in maritime industry. Process model development includes both analytical modeling and validation. Validation is done by comparison of the analytical results with those obtained by testing the models in the real operation conditions.

One of the projects in maritime application concerns the development of the special type of unconventional propulsion system. Its validation is based on comparison of the main characteristics obtained by the analytical model and results of the tests. This validation is done on the set of experimental points. Results of this validation will show the accuracy of the analytical model.

Another important aspect connected to the validation of the results of the tests is the development of the mathematical models based on the experimental results. The idea here is to create a data based model and introduce an extra feature of adjusting it to the new data.

The goal

The primal goal of this project is to create and test the models approximating/predicting data based model using Machine Learning Studio provided by Microsoft Azure. This model should be tested on several data sets to estimate the accuracy.

The second goal is to consider the possibility of introducing self learning models improving their accuracy based on new/ extra data.

To reach the goals of the project, it is required to address the following questions:

- Is it possible to build machine learning models to build the data based model in Machine Learning Studio?

- How to build such models?
- Which are the suitable methods?
- Is it possible to extend the concept to have a self-learning model adjusting automatically to produce a better fit while extra data are available?

Technical details

Model

The model is created in MATLAB and is represented as a nonlinear function of two parameters. The data provided are the sets of two input and one output points.

Experimental_points

The experimental set include 17 points in the format given above. These data will be provided in MATLAB or conventional text format.

Data

The analytical model could be used to generate extra data points (grids) to evaluate the accuracy of the created data based model. The data will be provided.

Contents of the project

1. Prestudy
   (a) Make a brief literature review to find the methods for approximating/fitting the data based model.
   (b) Get acquainted with the Machine Learning Studio in Microsoft Azure in general and machine learning algorithms in particular.
2. Work with data:
   (a) Load/import/export data
   (b) Data formats (csv or Excel)
   (c) Explore the ways of storing data on Azure
3. Create ML models
   (a) Make a data based model(s) using the predefined set points
   (b) Make it possible to evaluate the model on a bigger data set (grids)
4. Run experiments with different ML models

(a) Run experiments with models in 3a,b)

(b) Analyze and compare the results from different models.

(c) Think about the possibility of constructing self learning models

5. Reporting the results

(a) Describe the work done and results obtained;

(b) Prepare a "cleaned up" prototype and make sure that it is commented in a good way;

(c) Final report according to the course requirements and addressing the questions in the goal of the project.

Prerequisites

It is expected that students working on this project have experience in algorithm development, data and numerical analysis. Experience in Microsoft Azure is not required but good to have. Additionally, it might be beneficial to have a collaboration with the group working on the project "SCHEDULING ANALYSIS WITH MICROSOFT AZURE" to share the experience and help each other while exploring the Machine Learning Studio.

Working routine

Once a week, on Friday afternoon the whole group or each member himself should send a report (a simple Excel sheet) about the progress made. In the end of the course the developed prototype, report and a document describing the procedure of creation of the model and getting results should be handed in.