Algorithms and applications for machine learning based physical modelling

Bernhard Nornes Lotsberg

November 29, 2019

This report describes the master project for Bernhard Nornes Lotsberg, supervised by Simon Wolfgang Funke from the Scientific Computing group at the Simula Research Facility. The work will be performed from the spring of 2020 to the end of the spring of 2021.

The aim of this project is to investigate whether we can use machine learning algorithms to improve existing numerical models for physical phenomena. The model we will in particular investigate is the The Statkraft Hydrologic Forecasting Toolbox (Shyft) model (`https://gitlab.com/shyft-os`). This model is used for predicting the amount of water that will need to be stored in hydro power facilities. The problem is that it has lacking prediction power for some facilities described in the CAMELS data set [1]. The final goal of this project is to use additional meta data provided by Statkraft containing information about the environment around these facilities and apply Machine Learning to improve the prediction power of the model. If we manage to improve the model by use of Machine Learning, one could get an indicator of what parts of the meta data are important and this could hopefully give the developers of Shyft the ability to improve their physical modelling by taking into account this additional information.

1   Method

It is important to find out how to best implement a neural network to improve a physical model, and several configurations have to be tested. First off, the data is a time series, so we will employ a Recurrent Neural Network. Currently we have three potential ideas for how to train the network:

1. Train network on outcome data, using the physical model and the extra meta data as inputs.

2. Train a neural network on the entire system, including the original inputs of the physical model. Then learn difference between neural network and the original model, assuming the neural network does better than the model.

3. Same as point 2, but with some form of regularization using the physical model.

Most if not all code will be written in the Python programming language. For all these ideas and others, it is necessary to find out what kind of neural network that performs the best. Part of the thesis work will be to perform a literature study researching what neural networks function the best for problems like this.

2 Work plan

2.1 Fall 2019

1. STK-IN4300 - Statistical Learning Methods in Data Science. Theoretical introduction to machine learning.

2. FYS-STK4300 - Applied Data Analysis and Machine Learning. Hands-on introduction to machine learning.

3. FYS4150 - Computational Physics. Hand on introduction to numerical analysis.

2.2 Spring 2020

1. IN5400 - Image Analysis with Machine Learning. This course is highly relevant because of its high focus on deep learning.

2. IN4200 - High Performance Computing with Numerical Projects. This course is mandatory for the Computational Science master's program.

3. Do initial groundwork for thesis. Literature study to find promising neural network configurations. Identify ideas for simplified physical models to test neural network configurations on. Spend approximately two days per week at Simula, Fornebu for this.

2.3 Fall 2020

1. IN4110 - Problem Solving with High Level Languages. More Python coding is well suited for the thesis work. This course acts as a placeholder right now and can be switched out with a different one if necessary. Another possibility could be 10 credits special curriculum on Machine Learning applied on differential equations under Morten Hjorth-Jensen.

2. Apply promising neural networks to identified physical systems. Identify best way to train network.

3. If results from point 2 are promising, start applying the best learning method on the full Shyft model.

2.4 Spring 2021

1. Finish up research on machine learning applied on the Shyft model. Report performance, what went wrong and what can be improved.

2. Finish thesis by May 15th.

3 Milestones

Jun 2020 Have list of potential neural network configuration and simplified physical systems ready for implementation and evaluation.

Oct 2020 Have functioning neural network code. Have a well thought out physical system as test case.

Dec 2020 Have performance of potential neural network configurations measured, decide what to implement for Shyft if results are promising.

May 2021 Finish report.

3.1 Supervision

The main supervisor will be Simon Funke, and most research time will be spent at Simula to get supervision from him and his Phd. student Sebastian Mitusch.

3.2 Signature

Simon Wolfgang Funke:

References

[1] A. Newman, K. Sampson, M. Clark, A. Bock, R. Viger, and D. Blodgett, "A large-sample watershed-scale hydrometeorological dataset for the contiguous usa, boulder, co, ucar/ncar," 2014.