APL405
Machine Learning in Mechanics
Mid-Term Report

Machine learning for heat transfer correlations

Beomjin Kwon, Faizan Ejaz, Leslie K. Hwang

Submitted by:
Sukhbir Singh
2017ME10617

under the supervision of
Prof. Souvik Chakraborty
Assistant Professor, Department of Applied Mechanics
Indian Institute of Technology, Delhi

1.1 Data Generation

The initial and one of the most crucial phases of this project was the generation of data. This task consumed a lot of time since it involves updating of 3D model in SolidWorks and then updating the results of thermal simulation in ANSYS Fluent. After updating the results, I had to calculate the temperature at each rib and find out the heat transfer coefficients at each rib. After a week, I was left with some 50 data points which I think for the time being can be used for the ML part of the project. In the meanwhile, I will be extracting some more data points and will use all of them in the final report.

The data generation part involved the generation of a feature matrix consisting of the rib heights (number of ribs = 5) [e1, e2, e3, e4, e5] and the values of heat transfer coefficients at each rib [h1, h2, h3, h4, h5].

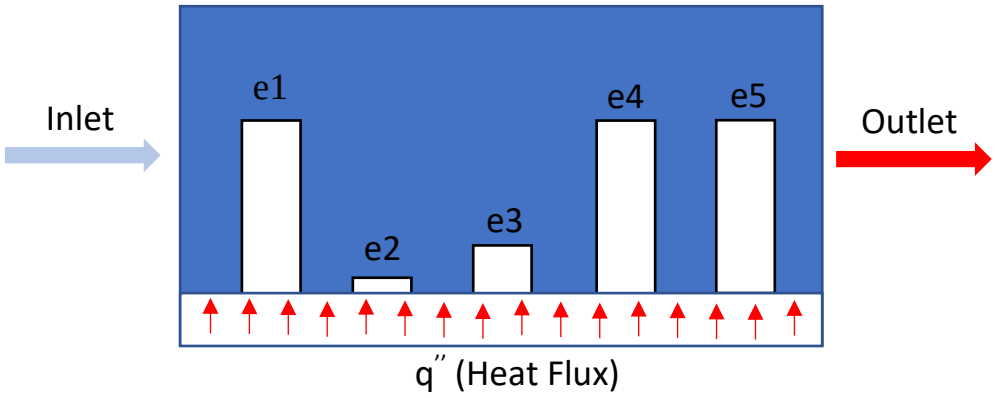

Fig.1. Schematic of channel with five ribs and heat flux

S. No.	Rib heights (mm)					Heat transfer coefficient (W/m2K)				
	e1	e2	e3	e4	e5	h1	h2	h3	h4	h5
1	1	0.0001	0.1	1	1	800	25	156	299	102
2	1	1	0.1	0.0001	1	832	31	99	202	149
3	1	0.0001	0.1	1	0.1	772	29	182	254	81
4	1	0.0001	0.1	1	0.0001	820	46	106	245	104
5	1	0.0001	0.1	0.1	1	773	30	168	241	97
6	1	0.0001	0.1	0.1	0.0001	763	20	136	226	107

Table 1. Sample data points collected from ANSYS Fluent Simulation

1.2 ML Model

The ML model I am trying to use in this project is the Random Forests Regression. RF algorithms are ensemble learning algorithms where it takes an average of the results or predictions obtained from a number of machine learning algorithms to get more precise and accurate results.

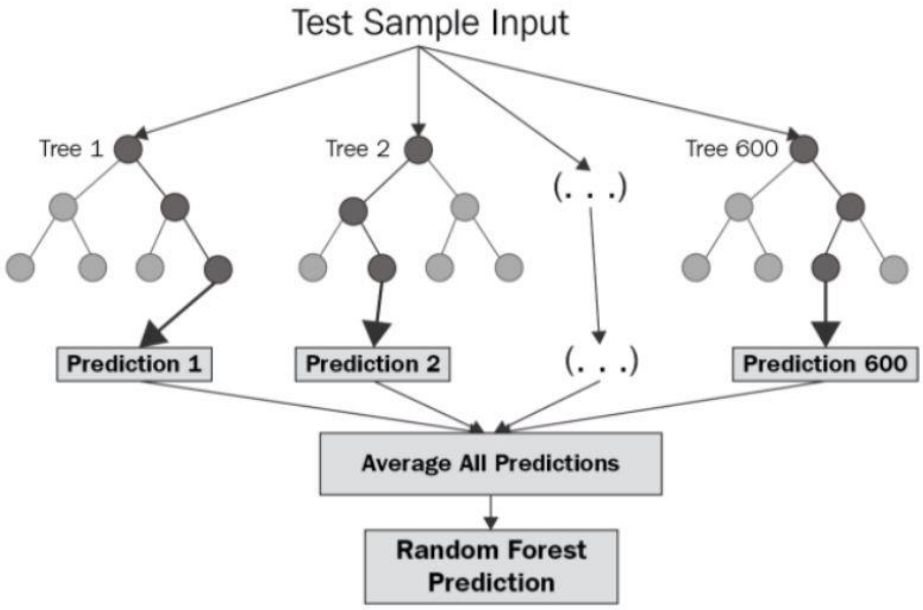

Fig.2. Structure of a random forest algorithm

The important hyperparameters used in these algorithms are:

1. Depth of the tree – max_depth
2. Seed number used by random number generator – random_state
3. Number of trees in the forest – n_estimator

According to the author of the paper being studied here, the most sensitive hyperparameter which affected the accuracy of the results was **n_estimator.** As we increase the number of decision trees working in the algorithm, the accuracy increases. The accuracy is defined by certain parameters like Mean Square Error or R^2 value. I have calculated both the values for the random forest algorithm as well as for the simple linear regression model as well to compare both the models and their accuracies.

1.3 Results

The python code that I wrote works for two models – one for the Random forest algorithm and the other for simple linear regressor. Out of the total number of data points, I selected 70% for the training and the remaining for the testing. The accuracy scores for both the models are given below:

Model	MSE (Mean Squared Error)	R^2 score
RF Algorithm	8.36	0.96
Simple Linear Regressor	23.91	0.58

We can compare the two models by either the MSE or R^2 value and it can be clearly seen that RF based model is better than simple linear regressor. And this difference of accuracies can be attributed to the nature of model used in RF algorithm. The averaging of results by RF algorithm reduces the error and makes the model a better alternative than simple linear regressor.

The plots of the two values (one with FVM model and other with ML model) compared side by side is shown below:

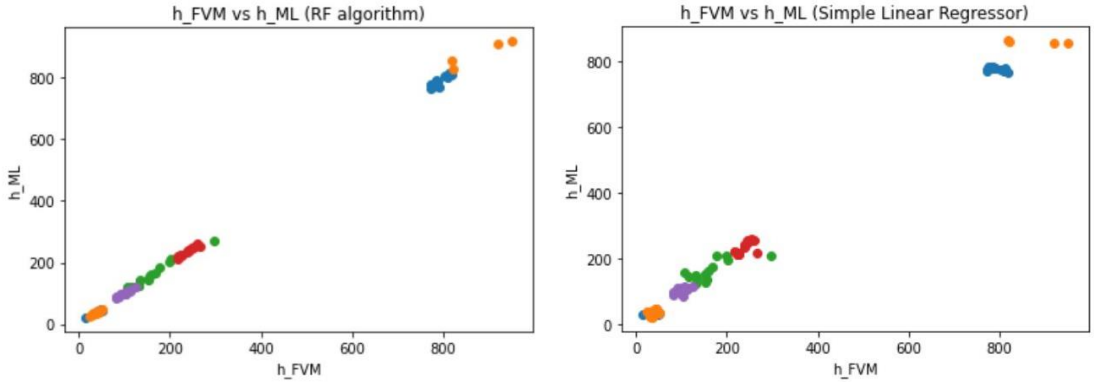

Fig.2. Comparison of heat transfer coefficients obtained by FVM (Numerical Methods) and ML models. The ML models used are RF algorithm and Simple Linear Regressor

It can be seen that in the model used by RF algorithm, the FVM values and the values obtained by ML are close to each other but in the one using simple linear regressor, it is more scattered showing that RF is better than simple regressor.

1.4 Conclusion and further work

In this report, I have focused on the comparison between the two algorithms which are used widely in the field of Machine Learning and how those two algorithms stand against each other. I used Python 3 for this task and different libraries like NumPy, Pandas and Scikit-Learn.

The next tasks in this project will include the variation of different hyperparameters in the RF algorithm for the purposes of optimization and calculating the R^2 values for heat transfer coefficients at different locations and for both, the training as well as testing data sets. Other task that I'm planning to take up is to test the interpolating capability of ML regressor. As you know that the rib heights were varied by only 3 values which are 1mm, 0.1mm and 0.0001mm. This time it can take any value and I will compare the results with the numerical model (using ANSYS) and see how well the ML regressor can take that.

1.5 Statement of contribution:

Sukhbir Singh – coded ML part, generated data, generated results, combined the data generation code with ML code.

1.6 References:

Fig. 2. - https://levelup.gitconnected.com/random-forest-regression-209c0f354c84