APL405
Machine Learning in Mechanics
Report 3

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. ML Model

The ML model I am using 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.

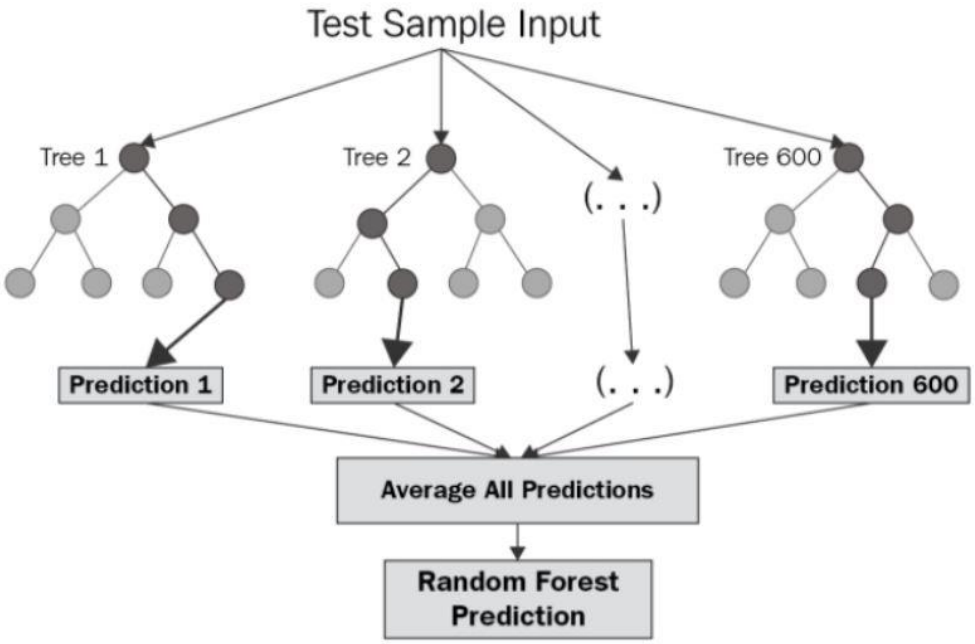


Fig.1. Structure of a random forest algorithm

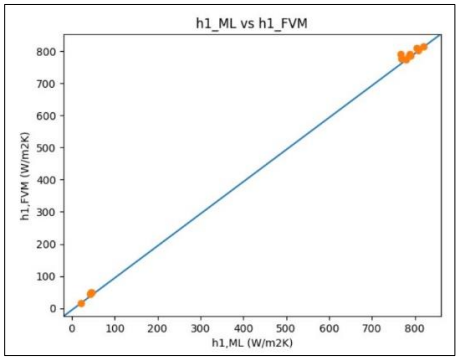
The important hyperparameters used in these algorithms are:

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_estimators

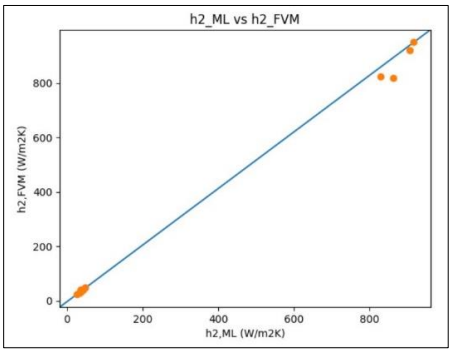
According to the author of the paper being studied here, the most sensitive hyperparameter which affected the accuracy of the results was **n_estimators.** At the end of this report, I have varied the value of n_estimator from 1 to 100 and tried to plot the accuracy (in terms of R^2 score) vs the n_estimator value.

2.1 Results – Comparison of ML and FVM model

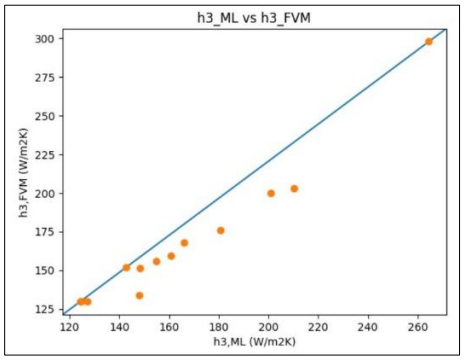
The FVM (Finite Volume model) gives us a numerical estimate for the heat transfer coefficient and its value was calculated using the ANSYS Fluent software. And from the data given by software, we tried to form a ML model which could give us accurate predictions for unknown rib geometries. The **Code1**[1] is a program which calculates the values of heat transfer coefficients at all the five locations or ribs and compares it with the values given by FVM model graphically. The output of the program can be seen from the graphs below.



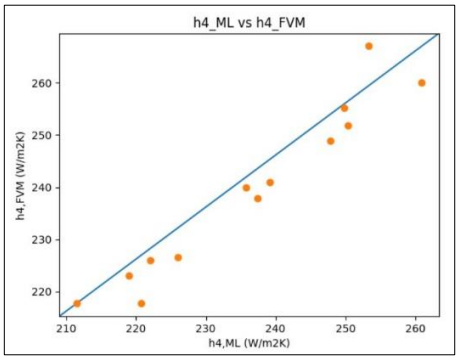
a.



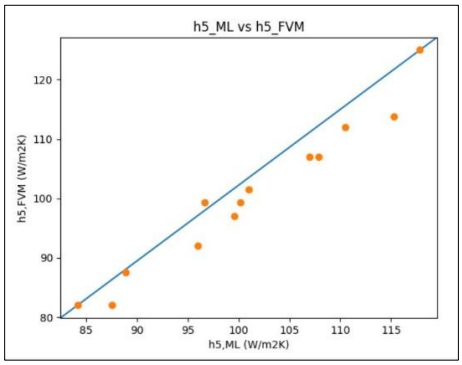
b.



c.



d.



e.

Fig. 2. Comparison of h predicted by FVM and ML regressor at (a.) first, (b.) second, (c.) third, (d.) fourth and (e.) fifth rib.

From the above graphs, it can be seen that for the upstream ribs, we are getting very accurate results but for the downstream ribs i.e. fourth and fifth ribs, the results deviate and that is because of the dependence of 'h' on the configuration of fluid flow at upstream ribs. There is a non-linear dependence on the flow configuration and the ML model cannot accurately predict the values at downstream ribs. This type of observation can be seen in the original research paper being studied.

2.2 Results – R^2 scores for training and testing sets

R^2 score tells us about how well can we explain the variance in the output or predicted values by the independent variable in a regression model. The formula for R^2 value can be written as:

$$R^2 = 1 - \frac{\text{Unexplained Variation}}{\text{Total Variation}}$$

Eqn. 1

The **Code2[2]** is the program which calculates the R^2 scores for the training and testing sets separately and for the 'h' values at all the rib locations. **The value of n_estimator used in the *Code2* is 10.** The results of the program can be seen in the below table.

R^2	h1	h2	h3	h4	h5
Training	0.993	0.993	0.957	0.958	0.942
Testing	0.999	0.997	0.964	0.765	0.933

Table 1. *R^2 score for testing and training sets*

The unexpectedly high value of R^2 values can be attributed to the low number of data points fed into the model for training. As it can be seen in the training set, the rib height of first rib is not varied much. Therefore, the R^2 values for first rib is very high as in the testing set also, there is not much variation in first rib height. For the next report, I'll vary the first rib height and update my data for a better evaluation and realistic results.

2.3 Results – R^2 score vs n_estimators in Random Forests Algorithm

As mentioned in the beginning of this report, the accuracy of random forests algorithm depends on the parameter **n_estimators.** It basically tells us how many decision trees are there in the algorithm. Since this algorithm gives us an average of all the predictions made by each tree, a greater number of trees will give better results. For this purpose, the **Code3**[3] gives us an idea about how increasing the value of n_estimators will give better results. In **Code3,** the h value we are using is h5 and we'll see how R^2 score for h5 varies with varying n_estimators value. The value of n_estimators is varied from 1 (single tree or simple linear regression) to 100. The output of the program is given below.

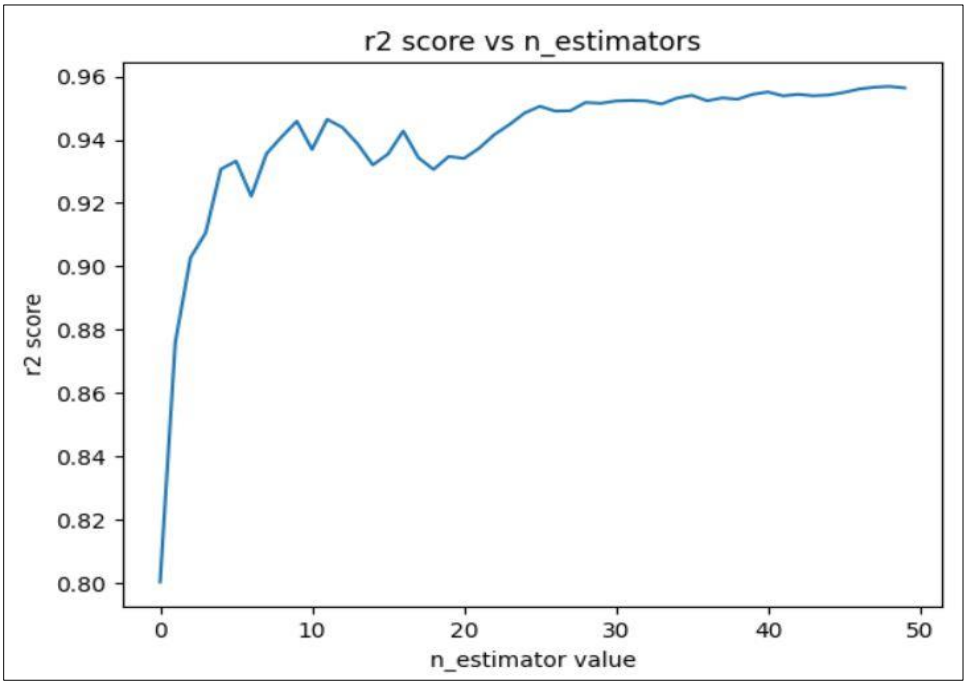


Fig. 3. R^2 score vs n_estimators value

3. Conclusion and further work

In this report, I have focused on the accuracy of the RF algorithm and how it can be improved by changing the number of trees used in the algorithm.

The next task in this project 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.

4. Statement of contribution:

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

5. Figures and Tables

1. **Fig.1.** Structure of a random forest algorithm – Explains how Random Forests Algorithm works to give a more accurate output than a simple linear regressor.

2. **Fig. 2.** Comparison of h predicted by FVM and ML regressor at (a.) first, (b.) second, (c.) third, (d.) fourth and (e.) fifth rib. – Graphs depicting the closeness of ML model to the FVM model where h values given by ML model corresponding to a particular rib geometry is compared to the value given by FVM model.

3. **Table 1.** R^2 score for testing and training sets – Table consisting of R^2 score which give us an idea about the accuracy of ML model at different rib locations.

4. **Fig. 3.** R^2 score vs n_estimators value – Graph depicting how the accuracy of RF regressor increases with number of trees used in the algorithm.

6. References:

1. Fig. 1. - https://levelup.gitconnected.com/random-forest-regression-209c0f354c84

2. Eqn. 1. - https://www.investopedia.com/terms/r/r-squared.asp#:~:text=R%2Dsquared%20(R2),variables%20in%20a%20regression%20model.&text=It%20may%20also%20be%20known%20as%20the%20coefficient%20of%20determination.