Prediction of particulate matter 2.5 using neural networks

Taehyuk Lee[1], Changhoon Lee[1,2*]

[1]Department of Mechanical Engineering, University of Yonsei, Seoul, Korea
[2]Department of Computational Science and Engineering, University of Yonsei, Seoul, Korea
*Corresponding author: clee@yonsei.ac.kr

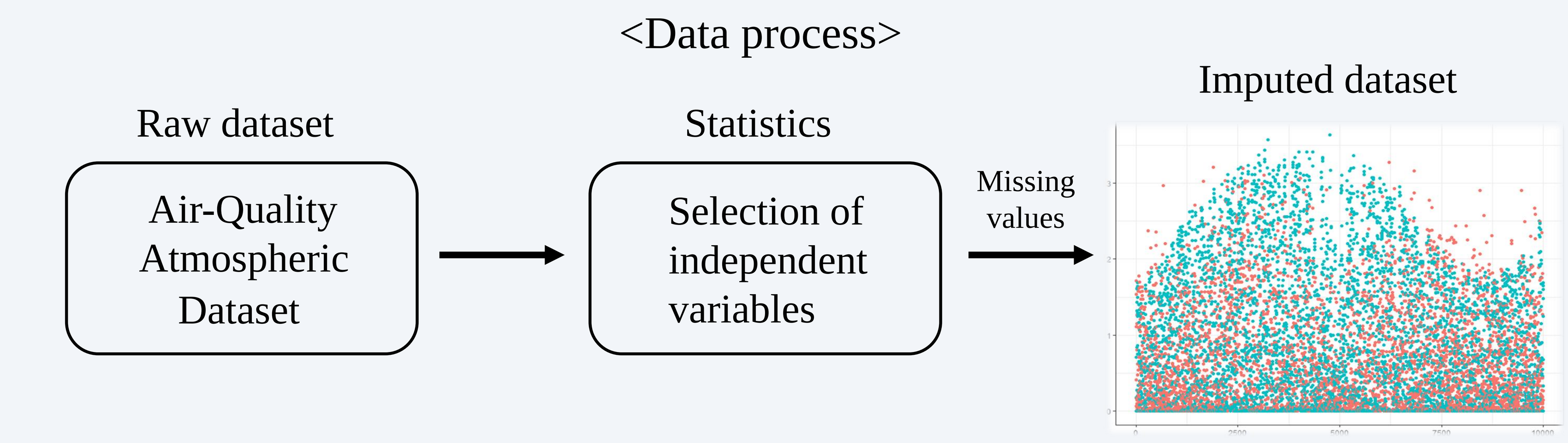
연세대학교
기계공학부

Introduction

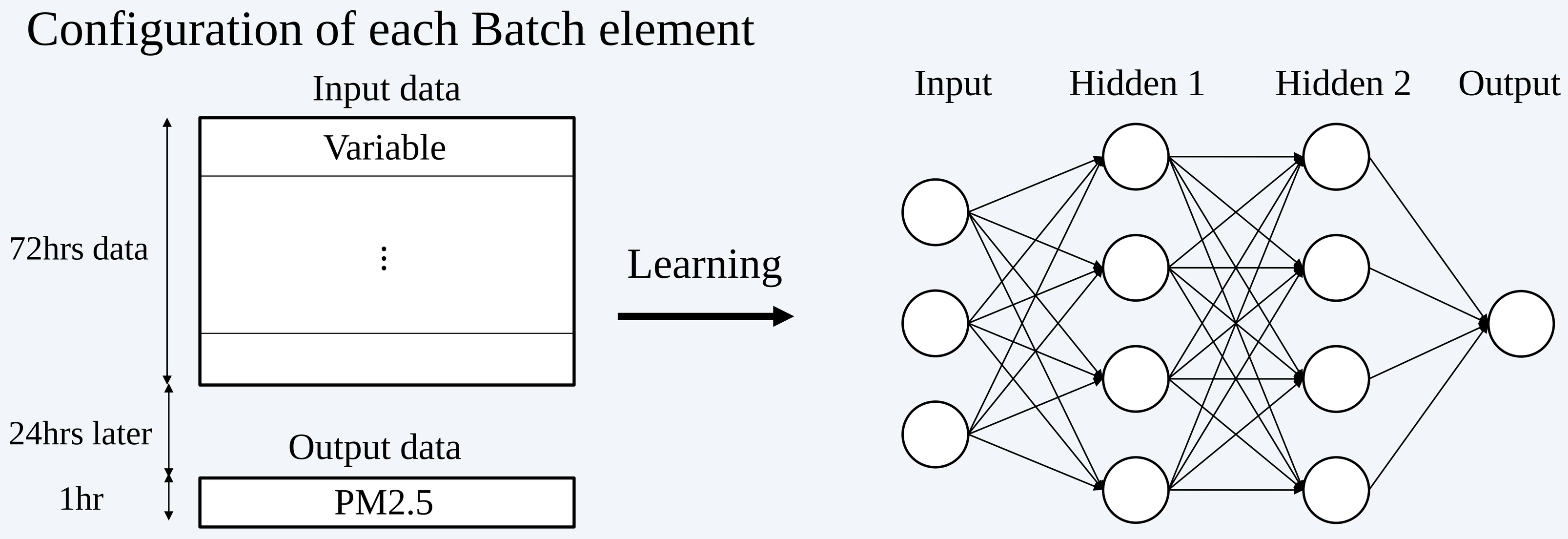
Recently, as fine dust has come to social problem, the efficiency and accuracy of dust predictive models have become an issue. Currently, most of predictive models are constructed in a way to solve multiple governing equations such as continuity equation, Thermodynamics 1st and 2nd law in a chain. **It requires a significant amount of computation and a funding**. In addition, the above governing equations **do not have topological information** and therefore it reveals **limitations**

In this study, in order to improve the limitations, we attempted to **identify and process the independent variables** related to the fine dust PM2.5 statistically among the various variables of the meteorological, air quality data. and **learned to the neural network**.

Schematic diagram



<Neural Network – Supervised learning>



Methods

◆ Data process

<Selection of variables for PM2.5>

a. Model (multiple linear regression)

$\hat{y}_i = b_0 + b_1 x_{1i} + b_2 x_{2i} \cdots + b_{ki} x_{ki}$

b. Akaike Information Criterion

$AIC = 2k - 2\ln(L)$

k: number of parameters
L: likelihood function

Stepwise selection method

<Selection of Independent variables>

c. Multicollinearity

$$VIF_i = \frac{1}{1 - R^2_i}$$

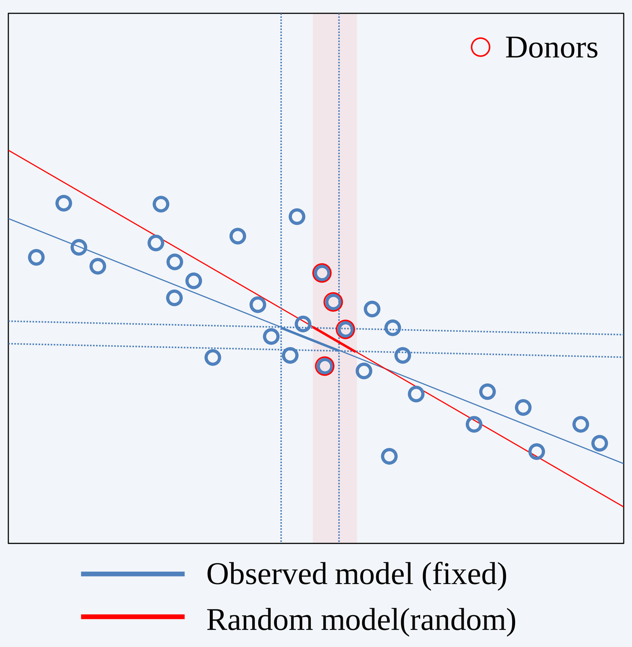
R^2_i: R squared

<Data Imputation>

d. pmm (predictive mean matching)

$\dot{y} = \hat{\beta}_0 + X_1 \hat{\beta}_1 + X_2 \hat{\beta}_2 + \cdots$

$\hat{\beta}_1, \hat{\beta}_2, \cdots, \hat{\beta}_n$ ~ Posterior Distribution

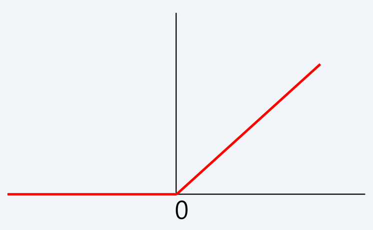


◆ Neural Network

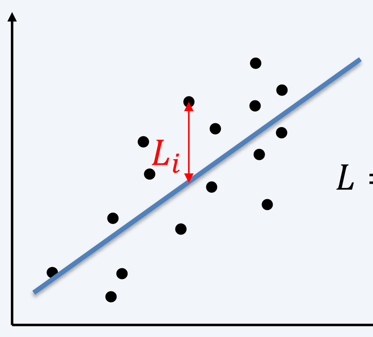
1. Neuron – One node

a. Linear term
Regression: f(w,b) = Wx_i +b

b. Activation function
Relu = max(0, z)



2. Loss function (Cost function)



$$L = \frac{1}{N}\sum_i (f(W,b) - y_{real})^2 + \lambda R(w)$$

3. Hyper parameter
Number of neuron: 256
Layer: 2

4. Optimizer (Learning method)
AdamOptimizer

5. Weight Initializer
Layer variance scaling initializer

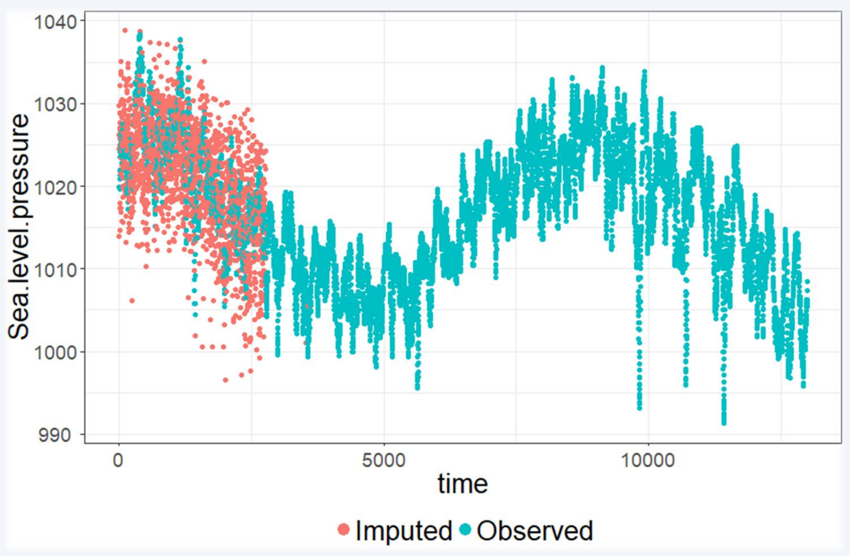
Result

◆ Selection of independent variables for PM2.5

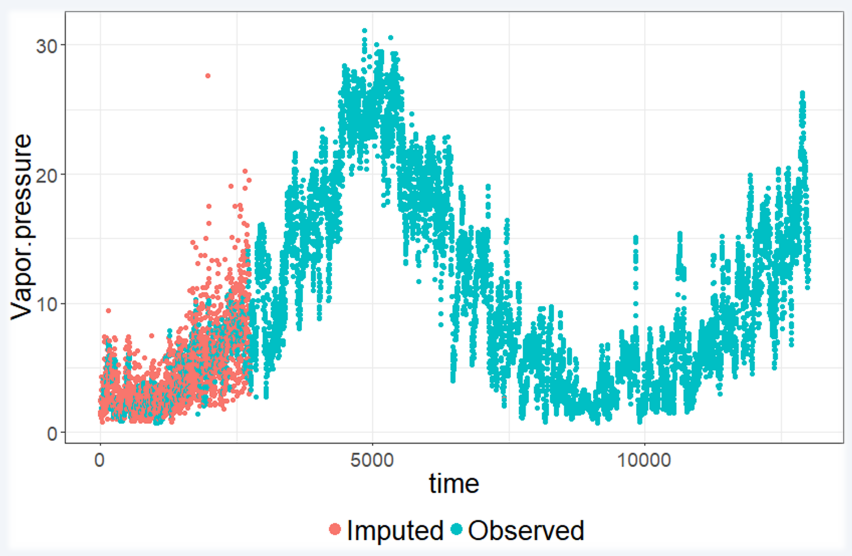
[Table . VIF for selected variables]

PM 10	Visibility	SO2	CO	NO2
1.847504	2.034791	1.567391	2.338489	2.776172
Vapor pressure	**O3**	**Sea level pressure**	**Temperature**	**Solar radiation**
6.484002	2.341847	2.991160	7.539861	2.048951
Total cloudy	**Wind direction**	**Wind speed**		
1.595853	1.292969	1.419134		

◆ Imputed dataset

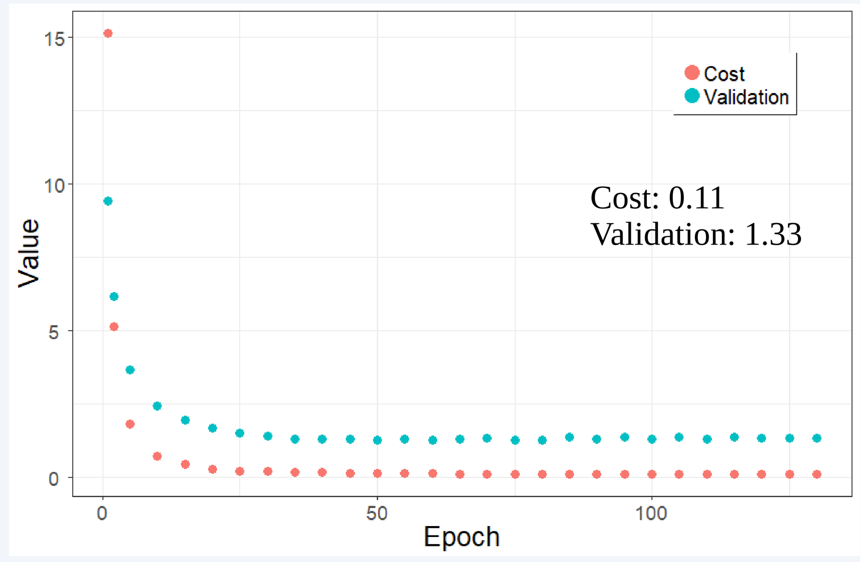


<Imputed data for Sea level pressure>



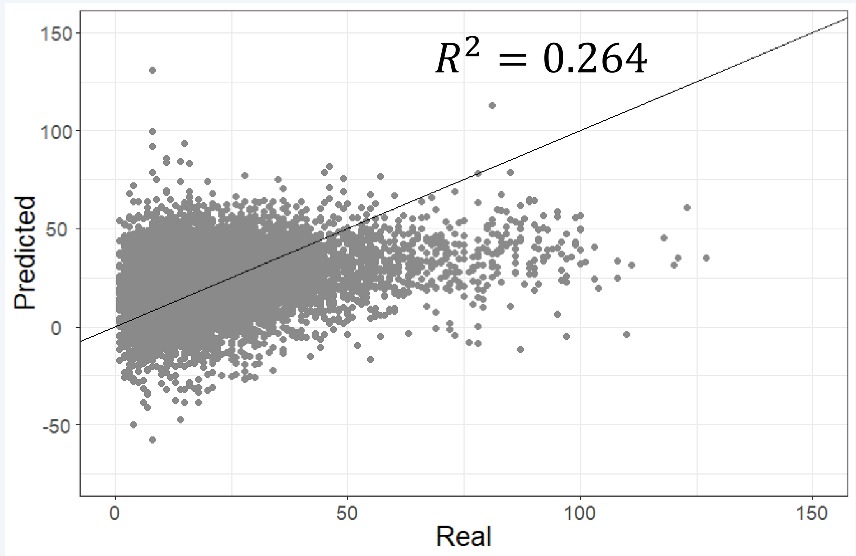
<Imputed data for Vapor pressure>

◆ Cost, Validation

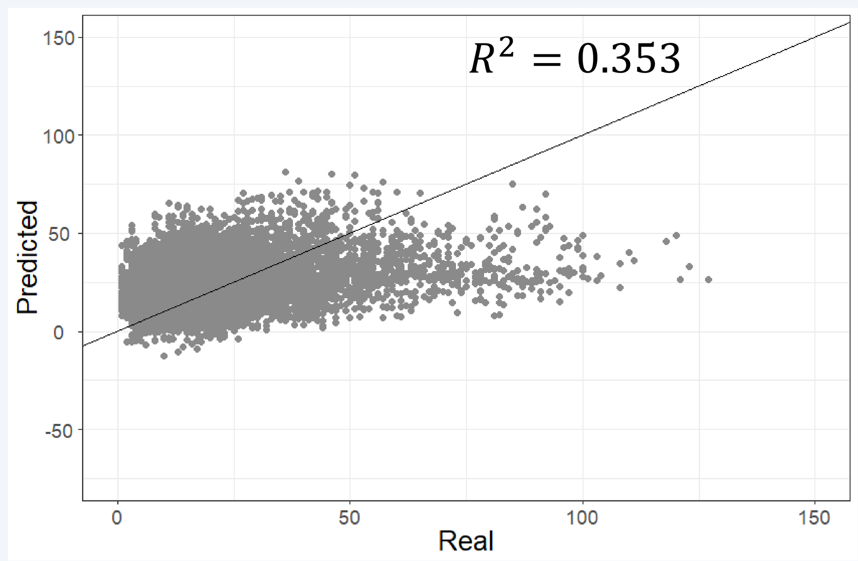


<Neural Network cost & validation>

◆ Accuracy (Scatter plot real & predicted value, R-squared)

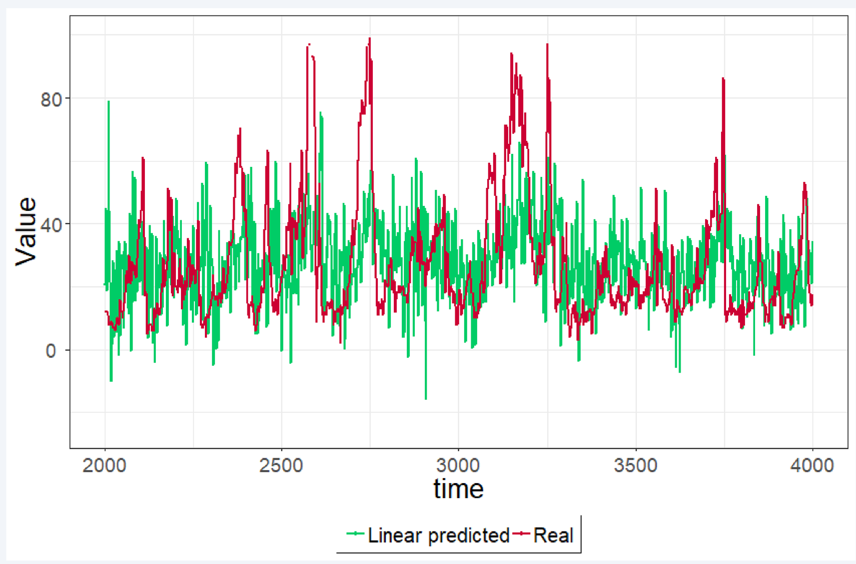


<Linear regression R-squared>

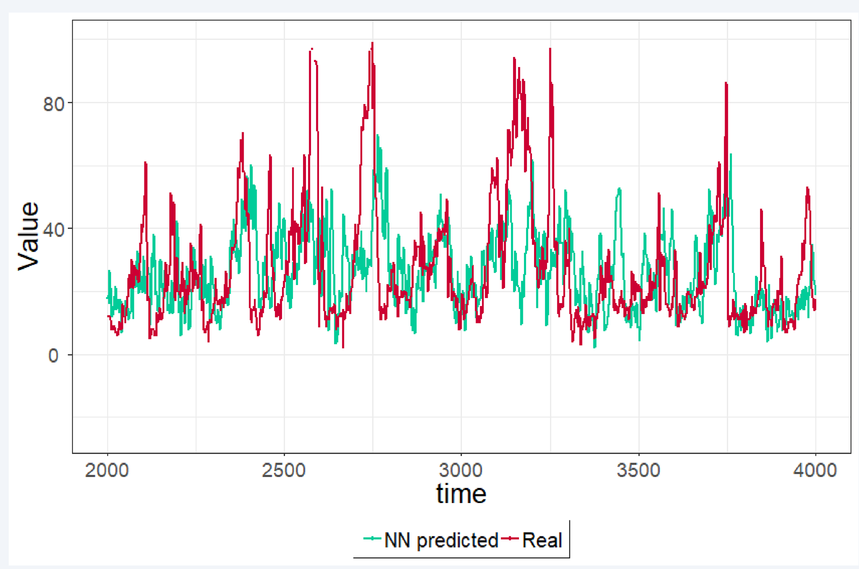


<Neural Network R-squared>

◆ Time variable



<Linear regression with time>



<Neural Network with time>

Conclusion

◆ Summary

In this research, we attempted to have Neural network learn refined dataset related to PM2.5 and obtained the predictive value for fine particulate matter. And then compared with predictions by multiple linear regression.

As a result, it was assured that the accuracy of the neural network is higher than the prediction result of multiple regression analysis by checking the distribution over time and real value.