Modeling Photocatalytic Degradation of Emerging Pollutants using a Neural - Evolutive Algorithm

Journal of Chemical Information and Modeling

Authors:

Izaguirre Gallegos, Tatiana L.; University Veracruzana, Faculty of Mechanical and Electrical Engineering, Poza Rica-Tuxpan.

Solis Maldonado, Carolina; Universidad Veracruzana Faculty of Chemical Sciences, Poza Rica-Tuxpan, Veracruz., Faculty of Sciences

Chemicals

Cristobal Salas, Alfredo; Veracruzana University, Faculty of Electronics and Communications Engineering, Poza Rica-Tuxpan.

Pérez Castañeda, Daniel; Veracruzana University, Faculty of Mechanical and Electrical Engineering, Poza Rica-Tuxpan.

Luna Sánchez, Raúl Alejandro; Veracruzana University, Faculty of Chemical Sciences, Poza Rica-Tuxpan.

Ortiz Silos, Nayeli ; Veracruzana University, Faculty of Chemical Sciences, Poza Rica-Tuxpan.

Barra Vázquez, Omar Alexander; Veracruzana University, Faculty of Electronics and Communications Engineering, Poza Rica-Tuxpan.

Cruz Miguel, Edson Eduardo; Veracruzana University, Faculty of Electronics and Communications Engineering, Poza Rica-Tuxpan.

García Martínez, José Román; Veracruzana University, Faculty of Electronics and Communications Engineering, Poza Rica-Tuxpan.

September 2024

INDEX

1. Install and configure tools
    a. Insta R

R is an open source programming language that provides software tools for data set manipulation, mathematical calculations, and graph visualization. The official site of the project can be consulted at the following link: https://www.r-project.org/. R is available for installation on various operating systems, including:

- Windows: Install R on Windows
- MacOS: Install R on MacOS
- Linux: Install R on Ubuntu

    b. Install Rstudio

On the other hand, Rstudio is a software that facilitates the management of the workspace for writing code, as well as the visualization and debugging of results. Likewise, Rstudio is available for the following operating systems:

- Windows: Install Rstudio on Windows
- MacOS: Install Rstudio on MacOS
- Linux: Install Rstudio on Ubuntu

2. Generate optimal model
    a. Generate figure 4 (Pareto Front)

The input file for this code is "**Experimental_data.csv**", which contains the experimental data of the degradation curves with 0.125 mg/L and 0.250 mg/L for 3, 5, 7 and 10 ppm.

i. Open the file "Figure_4".
ii. On line 28 modify the path to the file "**Experimental_data.csv**".
iii. Click on "File" and then "Save".
iv. Click on "Source".
v. In the "Plots" window you can view the Pareto front.
vi. The files that are generated at the end of the execution are:
    1. "**Optimal_model_weights.xml**": Contains in XML format the weights of the optimal neural network model.
    2. "**Optimal_model_data.csv**": Contains the number of layers, the architecture, the coefficient of determination (R^2) and the execution time (ms) of the optimal neural network model.
    3. "**Additional_models.csv**": Contains the same information as the file "Optimal_model_data.csv", but with respect to additional models that the algorithm suggests that meet the minimization objectives.
    4. "**Evaluated_models.csv**": Contains the same information as the file "Optimal_model_data.csv, but with data from all the models evaluated by the algorithm.
    5. "**Models_space.csv**": Contains the execution results of each evaluated model. This information contains the coefficient of determination (R^2), the execution time (ms) and a boolean value (0,1), where 1 indicates that it is a possible solution.
vii. Verify in the working directory that the output files have been created.
viii. Clean up the Rstudio workspace:
    1. Click on "Environment".
    2. Click on the "Cleaning" icon.

3. Click on "Plots".
4. Click on the “Cleaning” icon.

ix. Close the file “Figure_4”.

b. Generate figure 5 (NN Architecture)

The input file for this code is “**Experimental_data.csv**", which contains the experimental data of the degradation curves with 0.125 mg/L and 0.250 mg/L for 3, 5, 7 and 10 ppm. Additionally, the file “**Optimal_model_weights.xml”** which contains the weights of the optimal neural network model generated. Additionally, the file “**Optimal_model_data.csv**"which was generated in the code"Figure_4”. Therefore, these three files must exist in the same directory.

i. Open the file “Figure_5”.
ii. On line 22 modify the path to the file “**Experimental_data.csv**”.
iii. Click on “File” and then “Save”.
iv. Click on “Source”.
v. In the “Plots” window you can view the architecture with the weights of the optimal neural network model.
vi. Clean up the Rstudio workspace:
    1. Click on "Environment".
    2. Click on the “Cleaning” icon.
    3. Click on "Plots".
    4. Click on the “Cleaning” icon.
vii. Close the file “Figure_5”.

c. Generate Figure 6 (Model Correlation)

The input file for this code is “**Experimental_data.csv**", which contains the experimental data of the degradation curves with 0.125 mg/L and 0.250 mg/L for 3, 5, 7 and 10 ppm. Additionally, the file “**Optimal_model_weights.xml”** which contains the weights of the optimal neural network model generated. Additionally, the file “**Optimal_model_data.csv**"which was generated in the code"Figure_4”. Therefore, these three files must exist in the same directory.

i. Open the file “Figure_6”.
ii. On line 22 modify the path to the file “**Experimental_data.csv**”.
iii. Click on “File” and then on “Save”.
iv. Click on “Source”.
v. View the correlation data stored in the variable “correlacion”.
vi. In the “Plots” window you can view the graph that shows a linear regression.
vii. Clean up the Rstudio workspace:
    1. Click on "Environment".
    2. Click on the “Cleaning” icon.
    3. Click on "Plots".
    4. Click on the “Cleaning” icon.
viii. Close the file “Figure_6”.

3. Generate curves

a. Generate curve of figure 7 a) (3 ppm, 0.125 mg/L)

The input file for this code is “**Experimental_data.csv**", which contains the experimental data of the degradation curves with 0.125 mg/L and 0.250 mg/L for 3, 5, 7 and 10 ppm. Additionally, the file

“**Optimal_model_weights.xml”** which contains the weights of the optimal neural network model generated. Additionally, the file “**Optimal_model_data.csv**” which was generated in the code “Figure_4”. Therefore, these three files must exist in the same directory.

i. Open the file “Figure_7_a”.
ii. On line 23 modify the path to the file “**Experimental_data.csv**”.
iii. Click on “File” and then “Save”.
iv. Click on “Source”.
v. View the data of the variable “grafica_curvas” that contains the information on the degradation curves.
vi. Clean up the Rstudio workspace:
    1. Click on "Environment".
    2. Click on the “Cleaning” icon.
vii. Close the file “Figure_7_a”.

b. Generate curve from figure 7 b) (3 ppm, 0.250 mg/L)

The input file for this code is “**Experimental_data.csv**", which contains the experimental data of the degradation curves with 0.125 mg/L and 0.250 mg/L for 3, 5, 7 and 10 ppm. Additionally, the file “**Optimal_model_weights.xml”** which contains the weights of the optimal neural network model generated. Additionally, the file “**Optimal_model_data.csv**” which was generated in the code “Figure_4”. Therefore, these three files must exist in the same directory.

i. Open the file “Figure_7_b”.
ii. On line 23 modify the path to the file “**Experimental_data.csv**”.
iii. Click on “File” and then “Save”.
iv. Click on “Source”.
v. View the data of the variable “grafica_curvas” that contains the information on the degradation curves.
vi. Clean up the Rstudio workspace:
    1. Click on "Environment".
    2. Click on the “Cleaning” icon.
vii. Close the file “Figure_7_b”.

c. Generate curve from figure 7 c) (5 ppm, 0.125 mg/L)

The input file for this code is “**Experimental_data.csv**", which contains the experimental data of the degradation curves with 0.125 mg/L and 0.250 mg/L for 3, 5, 7 and 10 ppm. Additionally, the file “**Optimal_model_weights.xml”** which contains the weights of the optimal neural network model generated. Additionally, the file “**Optimal_model_data.csv**” which was generated in the code “Figure_4”. Therefore, these three files must exist in the same directory.

i. Open the file “Figure_7_c”.
ii. On line 23 modify the path to the file “**Experimental_data.csv**”.
iii. Click on “File” and then “Save”.
iv. Click on “Source”.
v. View the data of the variable “grafica_curvas” that contains the information on the degradation curves.
vi. Clean up the Rstudio workspace:
    1. Click on "Environment".

2. Click on the “Cleaning” icon.

vii. Close the file “Figure_7_c”.

d. Generate curve of figure 7 d) (5 ppm, 0.250 mg/L)

The input file for this code is “**Experimental_data.csv"**, which contains the experimental data of the degradation curves with 0.125 mg/L and 0.250 mg/L for 3, 5, 7 and 10 ppm. Additionally, the file “**Optimal_model_weights.xml”** which contains the weights of the optimal neural network model generated. Additionally, the file “**Optimal_model_data.csv**” which was generated in the code “Figure_4”. Therefore, these three files must exist in the same directory.

i. Open the file “Figure_7_d”.
ii. On line 23 modify the path to the file “**Experimental_data.csv**”.
iii. Click on “File” and then “Save”.
iv. Click on “Source”.
v. View the data of the variable “grafica_curvas” that contains the information on the degradation curves.
vi. Clean up the Rstudio workspace:
    1. Click on "Environment".
    2. Click on the “Cleaning” icon.
vii. Close the file “Figure_7_d”.

e. Generate curve of Figure 7 e) (7 ppm, 0.125 mg/L)

The input file for this code is “**Experimental_data.csv"**, which contains the experimental data of the degradation curves with 0.125 mg/L and 0.250 mg/L for 3, 5, 7 and 10 ppm. Additionally, the file “**Optimal_model_weights.xml”** which contains the weights of the optimal neural network model generated. Additionally, the file “**Optimal_model_data.csv**” which was generated in the code “Figure_4”. Therefore, these three files must exist in the same directory.

i. Open the file “Figure_7_e”.
ii. On line 23 modify the path to the file “**Experimental_data.csv**”.
iii. Click on “File” and then “Save”.
iv. Click on “Source”.
v. View the data of the variable “grafica_curvas” that contains the information on the degradation curves.
vi. Clean up the Rstudio workspace:
    1. Click on "Environment".
    2. Click on the “Cleaning” icon.
vii. Close the file “Figure_7_e”.

f. Generate curve from figure 7 f) (7 ppm, 0.250 mg/L)

The input file for this code is “**Experimental_data.csv"**, which contains the experimental data of the degradation curves with 0.125 mg/L and 0.250 mg/L for 3, 5, 7 and 10 ppm. Additionally, the file “**Optimal_model_weights.xml”** which contains the weights of the optimal neural network model generated. Additionally, the file “**Optimal_model_data.csv**” which was generated in the code “Figure_4”. Therefore, these three files must exist in the same directory.

i. Open the file “Figure_7_f”.
ii. On line 23 modify the path to the file “**Experimental_data.csv**”.

iii. Click on “File” and then “Save”.
iv. Click on “Source”.
v. View the data of the variable “grafica_curvas” that contains the information on the degradation curves.
vi. Clean up the Rstudio workspace:
    1. Click on "Environment".
    2. Click on the “Cleaning” icon.
vii. Close the file “Figure_7_f”.

g. Generate curve from figure 7 g) (10 ppm, 0.125 mg/L)

The input file for this code is “**Experimental_data.csv"**, which contains the experimental data of the degradation curves with 0.125 mg/L and 0.250 mg/L for 3, 5, 7 and 10 ppm. Additionally, the file “**Optimal_model_weights.xml”** which contains the weights of the optimal neural network model generated. Additionally, the file “**Optimal_model_data.csv**” which was generated in the code “Figure_4”. Therefore, these three files must exist in the same directory.

i. Open the file “Figure_7_g”.
ii. On line 23 modify the path to the file “**Experimental_data.csv**”.
iii. Click on “File” and then on “Save”.
iv. Click on “Source”.
v. View the data of the variable “grafica_curvas” that contains the information on the degradation curves.
vi. Clean up the Rstudio workspace:
    1. Click on "Environment".
    2. Click on the “Cleaning” icon.
vii. Close the file “Figure_7_g”

h. Generate curve of figure 7 h) (10 ppm, 0.250 mg/L)

The input file for this code is “**Experimental_data.csv"**, which contains the experimental data of the degradation curves with 0.125 mg/L and 0.250 mg/L for 3, 5, 7 and 10 ppm. Additionally, the file “**Optimal_model_weights.xml”** which contains the weights of the optimal neural network model generated. Additionally, the file “**Optimal_model_data.csv**” which was generated in the code “Figure_4”. Therefore, these three files must exist in the same directory.

i. Open the file “Figure_7_h”.
ii. On line 23 modify the path to the file “**Experimental_data.csv**”.
iii. Click on “File” and then “Save”.
iv. Click on “Source”.
v. View the data of the variable “grafica_curvas” that contains the information on the degradation curves.
vi. Clean up the Rstudio workspace:
    1. Click on "Environment".
    2. Click on the “Cleaning” icon.
vii. Close the file “Figure_7_h”

4. Final Comments

The file titled "Figure_generation.xlsx"includes the graphs generated from the data obtained from the codes"Figure_6.R", "Figure_7_a.R", "Figure_7_b.R", "Figure_7_c.R", "Figure_7_d.R", "Figure_7_e.R", "Figure_7_f.R", "Figure_7_g.R" and "Figure_7_h.R". Every eyelash of the file "Figure_generation.xlsx" is associated with a specific graph; for example, the "Figure_6" tab contains the output data of the code "Figure_6.R".

On the other hand, it is It is important to note that, as it is a model based on random numbers, there is the possibility that the algorithm converges in different neural network architectures. However, this architecture represents a version of the model that seeks to balance the established objectives. This implies that, as the population or the number of iterations increases, the probability of obtaining different results increases. In this context, since the solution space is large, there is a greater probability that one of these solutions is better. Therefore, the identification of these solutions requires considerable evaluation time, and since this model is heuristic, it will select one of the best evaluated architectures. The model converges by identifying a neural network architecture that efficiently balances the objectives to be minimized, which is a fundamental characteristic of it.