论文第五章中的提示工程-DHE

1. 生成使用随机森林方法处理离群值的脚本的DHE提示工程

You are an expert at data science. I have 550 pieces of data collected from the hydrocracking unit in a refinery, and I want to establish neural networks that use the properties of feedstock and operating conditions to predict the properties and impurities content of hydrocracking products. And then find the key impact factors of the output variables based on the models.

Your mission is to guide me to complete this process. Previously, you have introduced five stages of the process as below:

Stage 1: Data Preprocessing

Stage 2: Model Selection

Stage 3: Model Training and Optimization

Stage 4: Model Evaluation and Validation

Stage 5: Model Interpretation and Key Factor Identification

We are now at Stage 1. Previously, you have recommended some methods for handling outliers. I have chosen isolated forest method for my project. Now, generate the Python script for handling the outliers using isolated forest. The script should have the functions below:

1. Load data from excel file. The script should load the data from the Excel file in my folder.

2. Handling the outliers using isolated forest. I have uploaded my data to you. Analyze the dataset and recommend a set of parameters for handling the outliers.

3. Output data to Excel file. The script should output the results to an Excel file.

Should you have any question, ask me in your response.

*Attached file: Dataset.xlsx*

2. 生成训练神经网络模型的脚本的DHE提示工程

You are an expert at data science. I have 550 pieces of data collected from the hydrocracking unit in a refinery, and I want to establish neural networks that use the properties of feedstock and operating conditions to predict the properties and impurities content of hydrocracking products. And then find the key impact factors of the output variables based on the models.

Your mission is to guide me to complete this process. Previously, you have introduced five stages of the process as below:

Stage 1: Data Preprocessing

Stage 2: Model Selection

Stage 3: Model Training and Optimization

Stage 4: Model Evaluation and Validation

Stage 5: Model Interpretation and Key Factor Identification

We are now at Stage 3 and 4. 493 data remain after Stage 1. Previously, you have recommended some neural networks for me on Stage 2. I have chosen three models for my project: Convolutional Neural Network (CNN), Long Short-Term Memory Network (LSTM) and Backpropagation Neural Network (BP). Now, generate the MATLAB scripts for training, validating, optimizing and evaluating the three models. The description of the scripts is below:

1. The script for CNN. First, the script should load the data from the CSV file. The CSV file has 32 columns, the former 26 columns are input features and the last 6 columns are output features. Second, the script should divide the data into training set and testing set in a ratio of 8:2, and the training set and testing set should be saved in a CSV file. Third, the script should train the CNN model using the architecture and hyperparameters that you recommend, and optimize the hyperparameters automatically. The trained model should be saved in a MATLAB data file (Mdl\_CNN.mat). Fourth, the script should evaluate the model on the training set and testing set. The evaluation parameters should be calculated between the actual values and the predicted values of output features, the parameters are: Pearson correlation coefficient, mean absolute error, and root mean square error. The evaluation results should be saved in a CSV file.

2. The script for LSTM. The functions of this script are similar to the first script. The only difference is that on step 3, the script should train the LSTM model instead of CNN model, and the trained model should be saved in Mdl\_LSTM.mat.

3. The script for BP. The functions of this script are similar to the first script. The only difference is that on step 3, the script should train the BP model instead of CNN model, and the trained model should be saved in Mdl\_BP.mat.

Should you have any question, ask me in your response.

3. 生成使用Grad-CAM方法解释CNN的脚本的DHE提示工程

You are an expert at data science. I have 550 pieces of data collected from the hydrocracking unit in a refinery, and I want to establish neural networks that use the properties of feedstock and operating conditions to predict the properties and impurities content of hydrocracking products. And then find the key impact factors of the output variables based on the models.

Your mission is to guide me to complete this process. Previously, you have introduced five stages of the process as below:

Stage 1: Data Preprocessing

Stage 2: Model Selection

Stage 3: Model Training and Optimization

Stage 4: Model Evaluation and Validation

Stage 5: Model Interpretation and Key Factor Identification

We are now at Stage 5. Previously, 493 data remained after handling the outliers using the method you recommended on Stage 1, and I chose three neural networks based on your recommendation on Stage 2, and they were trained, optimized, validated and evaluated on Stage 3 and 4. You have introduced some methods for interpreting Convolutional Neural Network (CNN) to me. And I select Grad-CAM (Gradient-weighted Class Activation Mapping) method. Generate me a MATLAB script to execute the Grad-CAM explanation, the script should have the function below:

1. Load data. The script should load the training set that is stored in CSV file and the CNN model that is stored in MATLAB data file.

2. Execute Grad-CAM. The script should execute Grad-CAM explanation for all 6 output features. Every entry in the training set should be used as a query point.

3. Output data. The script should output the Grad-CAM results in a CSV file.