##

Title: Programs for 2D*J* prediction

Date of Update: 13/Nov./2020

Version: 1.0

Author: Kengo Ito[1] and Xiangru Xu[2]

Affiliation: [1]RIKEN Center for Sustainable Resource Science (CSRS), Environmental Metabolic Analysis Research team (EMAR), [2]Institut des Sciences et Ingénierie Chimiques, Ecole Polytechnique Fédérale de Lausanne (EPFL)

URL: http://dmar.riken.jp/NMRinformatics/

Reference: Ito, K., Xu, X., and Kikuchi, J. "Improved prediction of carbonless NMR spectra by the machine learning of theoretical and fragment descriptors for environmental mixture analysis" (submitted)

##

About

Programs for prediction of 2D *J*-resolved NMR spectroscopy by combination approach with quantum mechanical computation and machine learning, including the machine learning input file generator, predicted 2D *J*-resolved NMR spectra generator, and predicted *J*-coupling tree drawer, and so on.

Requirements

1. Python
   rdkit
   openbabel
   nmrglue
2. R
   caret
3. Java
4. Node.js
   openchemlib-extended
5. Gaussian

Example of using

(1) Quantum mechanical computation

[Data and programs: 1_QM directory in Predict_2DJ_programs_and_data.zip]

1-1. Execute command for DFT calculation

```
g16 33032.com
```

1-2. Convert Gaussian log files to sdf and mol files

```
cd your_file_path/1_QM/
```

python ob_conv.py your_file_path/1_QM

(2) The machine learning input file generation of $^{n}J_{HH}$ coupling constant for training dataset

[Data and programs: 2_JtrainDATA directory in Predict_2DJ_programs_and_data.zip]

2-1. Execute python program

cd your_file_path/2_JtrainDATA/

python Jtrain.py

2-2. GUI input arguments

① your_file_path/2_JtrainDATA/

② input_Jtrain.xlsx

③ your_file_path/1_QM/TrainData/log

④ your_file_path/1_QM/TrainData/sdf

⑤ your_file_path/1_QM/TrainData/mol

⑥ your_openchemlib-extended_path/node_modules

(3) The machine learning input file generation of ^{1}H chemical shift for training dataset

[Data and programs: 3_CStrainDATA directory in Predict_2DJ_programs_and_data.zip]

3-1. Execute python program

cd your_file_path/3_CStrainDATA/

python CStrain.py

3-2. GUI input arguments

① your_file_path/3_CStrainDATA/

② input_CStrain.xlsx

③ your_file_path/1_QM/TrainData/log

④ your_file_path/1_QM/TrainData/sdf

⑤ your_file_path/1_QM/TrainData/mol

⑥ your_openchemlib-extended_path/node_modules

(4) The machine learning input file generation of $^{n}J_{HH}$ coupling constant for test dataset

[Data and programs: 4_JtestDATA directory in Predict_2DJ_programs_and_data.zip]

4-1. Execute python program

cd your_file_path/4_JtestDATA/

python Jtest.py

4-2. GUI input arguments

① your_file_path/4_JtestDATA/

② input_Jtest.xlsx

③ your_file_path/1_QM/TestData/log

④ your_file_path/1_QM/TestData/sdf

⑤ your_file_path/1_QM/TestData/mol

⑥ your_openchemlib-extended_path/node_modules

(5) The machine learning input file generation of ^{1}H chemical shift for test dataset

[Data and programs: 5_CStestDATA directory in Predict_2DJ_programs_and_data.zip]

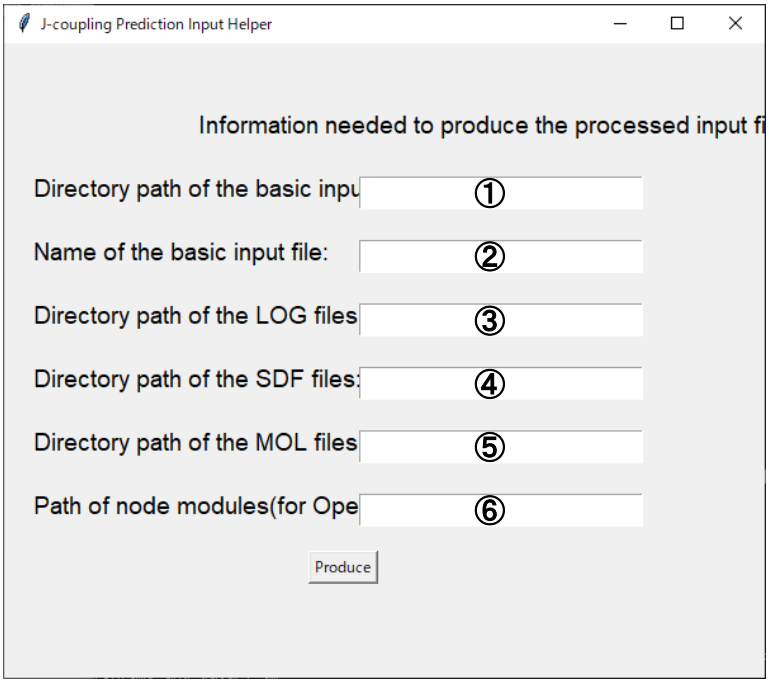
5-1. Execute python program

```
cd your_file_path/5_CStestDATA/
python CStest.py
```

5-2. GUI input arguments

① your_file_path/5_CStestDATA/
② input_CStest.xlsx
③ your_file_path/1_QM/TestData/log
④ your_file_path/1_QM/TestData/sdf
⑤ your_file_path/1_QM/TestData/mol
⑥ your_openchemlib-extended_path/node_modules

(2~5) Common GUI setting



(6) Generate fragment descriptors

[Data and programs: 6_GenFra directory in Predict_2DJ_programs_and_data.zip]

6-1. Execute batch program

```
cd your_file_path/6_GenFra/CS_train/water/
./toolgaussianlearndata.bat
cd your_file_path/6_GenFra/CS_train/methanol/
./toolgaussianlearndata.bat
cd your_file_path/6_GenFra/CS_test/water/
./toolgaussianlearndata.bat
```

More details are shown in previous study.

(7) Machine learning of $^{n}J_{HH}$ coupling constant

[Data and programs: 7_Jlearning directory in Predict_2DJ_programs_and_data.zip]

7-1. Execute R program

```
cd your_file_path/7_Jlearning/
Rscript ML_Pred_tool_CMD.R Core=2 K_CV=5 Grid=3
  ML_Numbers=36,48,145,169,173,191-192,236-237 outFile_Name=J_Train
  inFile_Path=your_file_path/7_Jlearning/ML_input_J_train.csv
```

(8) Machine learning of ^{1}H chemical shift

[Data and programs: 8_CSlearning directory in Predict_2DJ_programs_and_data.zip]

8-1. Execute R program

```
cd your_file_path/8_CSlearning/
Rscript ML_Pred_tool_CMD.R Core=2 K_CV=5 Grid=3
  ML_Numbers=36,48,145,169,173,191-192,236-237 outFile_Name=CS_Train
  inFile_Path=your_file_path/8_CSlearning/ML_input_CS_train.csv
```

(9) Prediction of $^{n}J_{HH}$ coupling constant

[Data and programs: 9_Jpred directory in Predict_2DJ_programs_and_data.zip]

9-1. Execute R program

```
cd your_file_path/9_Jpred/
Rscript ML_Test_tool_CMD.R Rdata_Path=your_file_path/7_Jlearning/J_Train.Rdata
  test_File_Path=your_file_path/9_Jpred/ML_input_J_test.csv test_Out_Name=J_Test
```

(10) Prediction of ^{1}H chemical shift

[Data and programs: 10_CSpred directory in Predict_2DJ_programs_and_data.zip]

10-1. Execute R program

```
cd your_file_path/10_CSpred/
Rscript ML_Test_tool_CMD.R Rdata_Path=your_file_path/8_CSlearning/CS_Train.Rdata
  test_File_Path=your_file_path/10_CSpred/ML_input_CS_test.csv test_Out_Name=CS_Test
```

(11) Predicted 2D *J*-resolved NMR spectra generation

[Data and programs: 11_2DJgen directory in Predict_2DJ_programs_and_data.zip]

11-1. Execute python program

```
cd your_file_path/11_2DJgen/
python plot2DJ.py CSpred.csv Jpred.csv your_file_path/11_2DJgen/999/pdata/
```

(12) Draw the predicted *J*-coupling tree

[Data and programs: 12_JtreeDraw directory in Predict_2DJ_programs_and_data.zip]

12-1. Execute python program

```
cd your_file_path/12_JtreeDraw/
python plotTree.py your_file_path/11_2DJgen/TreeTable.csv 1
```