Manual for Windows

Can He

1st January, 2020

GitHub: https://github.com/SanyHe/pyro_processor

E-mail: sanyhew1097618435@163.com

Throughout this manual, commands to be entered on the command line (like cmd in windows)are in blue and entries that are specific to the data set to be processed or code script to be modified are in red. The words you need to figure out in Glossary part are in green. The presence of >>> before a command indicates that the command is a python script, which means that it should be excuted by a python interpreter. Please check whether python version => 3.9.0 in your computer. if having no idea to do, see how to check the version and download a new one in Others part.

Table of Contents

1.Installation

The package can be downloaded from github at https://github.com/SanyHe/pyro_processor

The package layout would be :

```
$ chdir
~/pyro_processor
$ tree
.
|- LICENSE
|- README.md
|- model_training
|   |-- models
|   |-- pattern1_data.xlsx
|   |-- pattern2_data.xlsx
|   |-- pattern3_data.xlsx
|- processing_directory
|- Manual_for_Windows.pdf
|- Manual_for_Mac.pdf
```

```
$ chdir
~/pyro_processor/processing_directory
$ tree
.
|- requirements.txt
|- setup.bat
|- setup.sh
|- dataset
|- execution.py
|   |-- cpx_data.xlsx
|   |-- spl_data.xlsx
|- pyroxene_processor
|   |-- __init__.py
|   |-- exception.py
|   |-- global_variable.py
|   |-- module.py
|   |-- processor.py
|   |-- test
|   |-- predictor_model
|   |-- classfier_model
|   |-- predictor_model
|- vir_env_mac
|   |-- bin
|   |-- include
|   |-- lib
|   |-- pyvene.cfg
|- vir_env_windows
|   |-- Scripts
|   |-- Include
|   |-- Lib
|   |-- pyvene.ofg
```

Don't worry, just follow the subsequent steps in this manual, you'll know how to execute it.

2.Configuration

Method 1

Download all the dependencies to your system's python interpreter. Enter the following command in the command line (like cmd in Windows) in a processing directory (which itself can have any name), where setup.bat is located.

```
 chdir

~\pyro_processor\processing_directory

 setup.bat
```

Method 2

Use vir_env_windows, python's virtual environment, directly we create before.

```
cd vir_env_windows/Scripts

activate.bat

cd ..\..
```

or

Choose ./vir_env_windows/Scripts/python.exe as your project's interpreter in an IDE (like PyCharm).

Method 3

Create your own virtual environment and download the dependencies to your virtual environment python.

```
python -m venv vir_env_name

cd vir_env_name/Scripts
```

```
activate.bat

cd ..\..

setup.bat
```

3.Pre-Processing

A specific directory structure needs to be created within a processing directory (which itself can have any name) using exactly the naming standards below：

- ./dataset/cpx_data.xlsx

  a excel sheet stroes the cpx data for processing

- ./dataset/spl_data.xlsx

  a excel sheet stores the spl data for processing, which is not necessary depending on the pattern chosen

4.Processing

Method 1: execute in an interactive way (like terminal in Mac)

*Notice1：it is only allowed to execute the following commands under your processing directory(which can have any name), not other directories.

```
chdir

~/processing_directory

python
```

or

```
python3

>>> from pyroxene_processor.processor import process

>>> process(cpx_data="cpx_data.xlsx", pattern=1)
```

Method 2: execute with python script

*Notice2: the codes in execution.py can be modified in your way.

```
python execution.py
```

or

```
python3 execution.py
```

5.Package detail

This package “pyroxene_processor” will provide you with a function to use. Please import it in this way:

```
>>> from pyroxene_processor.processor import process
```

Then, you can call it with three parameters (cpx_data, spl_data, pattern)depending on the pattern you choose, the patterns’ detail are as follows:

Pattern Options	Data set	Description
1	cpx_data.xlsx	Based on the trained model, it will return the prediction results and classification results of polynomial regression, artificial neural network, extra tree and random forest with the given data set. All the models have been trained with the data set I we provided.
2	cpx_data.xlsx	Based on the trained model, it will return the prediction results and classification results of linear regression, polynomial regression, extra tree and random forest with the given data set. All the models have

		been trained with the data set II we provided.
3	cpx_data.xlsx spl_data.xlsx	Based on the trained model, it will return the prediction results and classification results of linear regression, extra tree with the given data set. All the models have been trained with the data set III we provided.

*data set I: The whole data;

data set II: The data without those from clinopyroxene and omphacitite in spinel lherzolite

data set III: The data from clinopyroxene in spinel lherzolite

When you want to use pattern I to process, cpx_data.xlsx is needed.

```
>>> process(cpx_data="cpx_data.xlsx", pattern=1)
```

When you want to use pattern II to process, cpx_data.xlsx is needed.

```
>>> process(cpx_data="cpx_data.xlsx", pattern=2)
```

When you want to use pattern III to process, cpx_data.xlsx and spl_data.xlsx are needed.

```
>>> process(cpx_data="cpx_data.xlsx", spl_data="spl_data.xlsx",  pattern=3)
```

6.Glossary

pyroxene_processor: a python package provides the function "process", which is the combination of calculator, predictor and classifier.

vir_env_windows: on Windows a python’s virtual environment with all the dependencies we use when creating the pacakage. You can run the commands and execution.py when activate this virtual environment without running setup.bat.

dataset: this directory stores the data to process, please name the sheet "cpx_data.xlsx" if it is cpx data or name the sheet "spl_data.xlsx" if it is spl data.

requirments.txt: it's a text where all the dependencies we used when creating the package are listed. It is used for configuration. You can ignore it if you run setup.sh in the command line. Or you can run pip install -r requirements.txt in the command line.

excecution.py: it's a python scripts used to run the commands for processing the given data once you execute it in the command line. It is editable according to the routine we mentioned in Processing part.

results: this directory will be created automatically as soon as you run the command, such as >>>process(cpx_data="cpx_data.xlsx", spl_data="spl_data.xlsx", pattern=3) . It will store the calculation results, prediction results and classification results.

setup.bat: on Windows, a batch script, which can be executed directly in the command line, stores the command to download all the dependencies automatically we used when creating the package.

virtual environment: A virtual environment is a tool that helps to keep dependencies required by different projects separate by creating isolated python virtual environments for them.

7.Others

1.Python Download for Windows：

https://www.python.org/downloads/windows/

Please download the version 3.9.0.

- Download Windows x86-64 executable installer

2. Python and Python3:

When entering the following command in command line

Python

Or

Python3

It will open a python shell, which is an interactive way to run python command, and the information about which python version you are using. It’s better to check the version by those two commands to see which one you can use when to run execution.py.

Enter the following command will quit from the current python shell

```
>>> quit()
```

3. Difference between pip and pip3:

Pip is for python version less than 3. and **pip3** is used when you want to install packages for python version 3 or higher.