

Machine learning model identifies DNA methylation diversity in iPSCs as a distinguishing property of iPSCs from human ESCs

Source codes for generating machine learning models in this paper.

1. Description

1.1. learning.py

Source code written with Python3 which used for generating machine learning models in our paper.

1.2. query.py

Source code written with Python3 which used for getting classification results in our paper.

1.3. save.py

Source code written with Python3 which used for saving machine learning models in our paper.

1.4. json2txt.py

Source code written with Python3 which used for transforming JSON file to TXT file in our paper. The JSON file has the weight information of generated machine learning models.

1.5. make_sh.py

Source code written with Python3 which used for making following shell script files. When you run this script, several arguments have to be needed; details are written in below "Usage" section.

1.5.1. learning_30_query_save.sh

Job submission shell script used in our study. This script call "learning.py", "query.py", "save.py", "dump.sh".

1.5.2. dump.sh

Call jubadump method and get JSON file which has the weight information of generated machine learning models.

1.6. classifier1.00.json

JSON file which described hyper-parameter of machine learning.

1.7. config_learning.txt

Configuration file which described the file path and category of the class for using to generate machine learning models.

1.8. config_query.txt

Configuration file which described the file path and actual category of the class for using to classify the cell types by querying to generated machine learning models.

1.9. Dump/

Directory storing files generated by dump.sh.

1.10. LearningModel/

Directory storing files generated by save.py.

1.11. result.txt

Classificated results file geneated by query.py.

1.12. log.txt

Learning status log file.

2. Requirement

- jubatus (<http://jubat.us/ja/quickstart.html>)
- multiprocessing (<https://docs.python.org/ja/3/library/multiprocessing.html>)

3. Usage

1. Make config_learning.txt & config_query.txt

column number	description
1	Category of the class.
2	File path of DNA methylation data.

2. Start up jubatus searver.

```
1 | jubaclassifier -f classifier1.00.json
```

3. Run "make_sh.py" with argument described in below.
total_num (int) : number of total learning iteration. queryANDsave_num (int): period of each learning iteration for classification and saving model.

```
1 # python make_sh.py total_num queryANDsave_num
2 python make_sh.py 100 10
```