iterb-PPse: Identification of transcriptional terminators in bacterial by incorporating nucleotide properties into PseKNC

Description: The “iterb-PPse” is the method we proposed to predict terminators. B illustrates the prediction process of “iterb-PPse”. It extracts three features from gene sequences at first, namely Pse5NC-Ⅰ, Pse5NC-Ⅱ, 47 nucleotide properties. Then sort all features using F-score and select the best feature set by IFS. Finally utilizes trained XGBoost to determine whether these sequences are terminators.



1. **Introduction of the software**

The “iterb-PPse.zip” is the prediction tool we developed for terminators used C#. Just double click steup.exe in folder “Release”, you can install our software.

**Please note that it is best not to install the software under the C drive, otherwise you may encounter permission problems.**

**If you encounter problems during installation or use, you can send an email to 1533937610@qq.com, we will try our best to help.**

1. **Introduction of all codes**

|  |  |
| --- | --- |
| File | Description |
| dataprocess.m | Import DNA sequence from csv file into matlab and remove redundant information |
| physical properties.m | Extracting nucleotide properties from sequences |
| Pesknc12.m | Extracting PsekNC-Ⅰfrom sequences  Extracting PsekNC-Ⅱ from sequences |
| write4libsvm.m | Transform the data into the format needed for feature sorting  Input format is labeled feature matrix (label in last column)  Output is a txt file in a specific format |
| fselect.py | Feature ranking using F-score  Run the following statement on the command line: ‘ptyoh fselect.py data.txt’, data.txt represents the file obtained from write4libsvm.m |
| txTOm.m | Extract the permuted feature order from the output file “\*.fscore” |
| ifs.m | Incremental feature selection |
| featureSelect | feature selection |
| XGBoost.py | model XGBoost |
| crossvalid.m | 100 times 5-fold cross-validation |
| plot\_roc.m | plot ROC curve |