Cell Image Feature Calculation Tool User Manual 2

Credit by Yu Ishikawa-Yamauchi and Takeshi Nagata

Address: [ishikawa.yu.ym@yokohama-cu.ac.jp](mailto:ishikawa.yu.ym@yokohama-cu.ac.jp), [nagata.takeshi.ge@u.tsukuba.ac.jp](mailto:nagata.takeshi.ge@u.tsukuba.ac.jp)

■Overview

This tool reads images, calculates three types of image features, and performs learning and prediction of linear multiple regression.

■Operating Environment

This tool runs on Windows (64bit).

To use this tool, you need to install Matlab2022b runtime. Free installation is available.

Please download and install Matlab2022b Runtime from the following link.

<https://ssd.mathworks.com/supportfiles/downloads/R2022b/Release/0/deployment_files/installer/complete/win64/MATLAB_Runtime_R2022b_win64.zip>

■Startup

3

Double-click LSR.exe to start.

グラフィカル ユーザー インターフェイス

自動的に生成された説明

1. Specifies an image folder, for tif images only. All subfolders under the specified folder will also be searched.
2. The specified folder will be displayed.
3. Fractal dimension parameter: Specify the distance.
4. Fractal dimension parameter: Specify the neighborhood pixel range.
5. Specifies the offset of the analysis range.
6. Check this box if you want to visualize images side by side by likelihood value.
7. Specifies the string contained in the file name for Class 1 (Aged).
8. Perform learning.
9. When predicting: Select the learned model file (model.mat).
10. The model file name will be displayed.
11. Run the prediction.

■Smoothing

Smoothing is performed before calculating the feature values to reduce the influence of minute geometry.

(Since the optimal setting value is different for each feature, it cannot be changed from the GUI.)

This function uses a bilateral filter (Matlab function imbilatfilt).

https://jp.mathworks.com/help/images/ref/imbilatfilt.html

We use the following command.

img2 = imbilatfilt( img, SmoothingRate\*255^2, Distance )

■ Analysis Range

It is possible to specify the calculation range (offset) of the fractal dimension as follows.

The unit for specifying the offset is pixels.

■Learning young and aged using linear multiple regression

This function uses linear multiple regression to learn the following variables. Generally, in machine learning, the more complex the model, the higher the accuracy, but the explanatory nature of the internal calculations and results cannot be guaranteed. This tool uses a linear multiple regression model, placing more emphasis on the explanatory nature of the calculations and results than on accuracy.

・The explanatory variables x; three types of features (histogram of fractal dimension, histogram of vacancy area, histogram of edge intensity) are standardized (mean value subtracted and divided by standard deviation).

・The objective variable y; the objective variable y is set to 0 for young and 1 for aged. (Data with "A" in the file name are assumed to be "aged" and all others are assumed to be "young".)

The linear multiple regression equation is as follows, where β is the weight obtained by machine learning.

y

Ideally, the estimated value of y should be 0or1, but in practice, y can be less than 0 or greater than 1, so it is converted to likelihood p by the following sigmoid function. Since the range of p in this equation is [0 1], we can interpret p as the likelihood of aged.

p

■Visualization of calculation results

It is possible to visualize and check the images side by side for each value of likelihood (likelihood) p.

■Executing Learning

When the "Learn" button is pressed, a dialog box for selecting a folder for output opens, and when the folder is specified, the learning process begins. When the calculation is completed, the following file is output.

1. シャツ が含まれている画像

   自動的に生成された説明テーブル

   自動的に生成された説明 2) 3) 4) 5)　 6)

1)This is a learned model file. Required for prediction.

2) The output shows the features of the input data, the weights of the training model and the predictions of the teacher data.

3, 4) A graphical representation of the prediction results for the supervised data.

5) Input data sorted and arranged by likelihood.

6) Binarized results are output when the number of vacancies is calculated.

■Contents of result.csv

The learned weights and prediction results are output.

■About the graphical display of the prediction results

The horizontal axis takes the correct answer (0:young 1:aged) and the vertical axis plots the predicted likelihood.

■LOO (Leave One Out) cross-validation

In general, in machine learning, it is a matter of course that the training model reproduces the teacher data with high accuracy. What is important is the accuracy for unknown data not used for training. In the LOO (Leave One Out) cross-validation, the following validation is performed N times (N is the number of teacher data). From the N teacher data, one data is excluded, and a training model is created with N-1 teacher data. The excluded one data is input into the learning model to predict the objective variable. This process is performed N times to obtain the estimation results for all data. All these estimation results were not used to create the training model, so the predictive accuracy of the LOO cross-test can be regarded as the predictive accuracy for unknown data.

テキスト, テーブル

自動的に生成された説明■Prediction

(1) Specify the model.mat created in the study.

(2) Click the "Predict" button to open the folder selection dialog to select the folder to output the prediction results. A folder selection dialog opens.

When a folder is specified, the prediction will start, and the following file will be output to the specified folder.