

Description

Details

URL: www.devchakraborty.com

Abbreviations

- (F)ROC: (free-response) receiver operating characteristic
- ROI: region-of-interest
- AUC: $\int_0^1 \hat{A}(\hat{F})d\hat{F}$: area under the ROC curve
- FOM: figure of merit
- FP: false positive
- TP: true positive
- NL: non-lesion localization
- LL: lesion localization
- CI: confidence interval
- SP: specificity
- SE: sensitivity

Á
ÖÓT PÁMÁ
Á
UÜPÁMÁ

Dataset

The dataset is an external data file. The dataset is a list of 9 elements. The first element is a four-dimensional numerical array that contains all non-lesion localization ratings. The dimensions of the array are number of modalities * number of readers * number of all cases * number of non-lesion localization marks in a case. Assign -Inf to the ratings of the non-lesion localization that does not exist. Note that this array **must** have four dimensions. Assign one to the fourth dimension for the condition that the maximum non-lesion localization marks is one. For ROC datasets, false positive ratings are assigned to **NL**.

Data structure

The structure of the dataset used in analysis functions is an R list containing 9 elements. Following are the elements in the list.

- NL**: a four dimensions numerical array that contains all non-lesion localization ratings. The dimensions of the array are number of modalities * number of readers * number of all cases * number of non-lesion localization marks in a case. Assign -Inf to the ratings of the non-lesion localization that does not exist. Note that this array **must** have four dimensions. Assign one to the fourth dimension for the condition that the maximum non-lesion localization marks is one. For ROC datasets, false positive ratings are assigned to **NL**.
- LL**: similar with **NL** except its dimension **only** applies to number of abnormal cases and lesion localization. For ROC datasets, true positive ratings are assigned to **LL**.
- lesionNum**: an integer vector with length of number of abnormal cases indicating the number of lesions in each abnormal cases.
- lesionID**: an integer matrix with dimension of number of abnormal cases by maximum of **lesionNum**. It gives the ID information of lesions in each abnormal cases. For the cases with lesions less than the maximum lesion number, -Inf are assigned to the unavailable elements. **Note** that ratings of lesions in **LL** must appear as the sequence of **lesionID** for corresponding case. For example, if the **lesionID** for the first abnormal case is c(4, 2, 3, 1), the ratings in **LL** for this case must appear as this sequence.
- lesionWeight**: a numerical matrix with same dimension of **lesionID**, which is the relative importance of detecting each lesion. The weights must sum to unity for each case.
- maxNumMarks**: the maximum number of non-lesion localization.
- dataType**: the type of the dataset. It can be "ROC", "ROI" or "FROC".
- modalityID**: a vector of strings, which indicates the ID of each modality that is used in the dataset.
- readerID**: a vector of strings, which contains the ID of each reader. **Note** that the order of elements in **modalityID** and **readerID** must identical with **NL** and **LL**. For example, NL[1, 2, ,] indicates the ratings of the reader with the second ID in **readerID** using the modality with the first ID in **modalityID**.

Data file format

The package is able to read JAFROC, OR-DBM MPMC (ROC data only) and iMPMC (ROC data only) data files. They can be imported by using functions [ReadDataFile](#).

JAFROC data file format

The JAFROC data file is an Excel file containing three worksheets (*.xls and *.xlsx are supported): (1) `Truth` (2) `TP` or lesion localization worksheet and (3) the `FP` or non-lesion localization worksheet. Except for the `Truth` worksheet, where each case must occur at least once, the number of rows in the other worksheets is variable.

1. `Truth` worksheet consists of

- `CaseID`, an integer field uniquely labeling the cases (images). It must occur at least once for each case, and since a case may have multiple lesions, it can occur multiple times, once for each lesion.
- `LesionID`, an integer field uniquely labeling the lesions in each case. This field is zero for normal cases.
- `Weight`, a floating point field, which is the relative importance of detecting each lesion. This field is zero for normal cases and for equally weighted lesions; otherwise the weights must sum to unity for each case. This field is irrelevant unless a weighted figure of merit is selected, see next section.

2. `TP` worksheet consists of

- `ReaderID`, a string field uniquely labeling the readers (radiologists).
- `ModalityID`, a string field uniquely labeling the modalities.
- `CaseID`, see `Truth` worksheet. A normal case in this field will generate an error.
- `LesionID`, see `Truth` worksheet. An entry in this field that does not appear in the `Truth` worksheet will generate an error. It is the user's responsibility to ensure that the entries in the `Truth` and `TP` worksheets correspond to the same physical lesions.
- `TP_Rating`, a positive floating-point field denoting the rating assigned to a particular lesion-localization mark, with higher numbers represent greater confidence that the location is actually a lesion.

3. `FP` worksheet consists of

- `ReaderID`, see `TP` worksheet.
- `ModalityID`, see `TP` worksheet.
- `CaseID`, see `TP` worksheet.
- `FP_Rating`, a positive floating-point field denoting the rating assigned to a particular lesion-localization mark, with higher numbers represent greater confidence that the location is actually a lesion.

Stacked input format for OR-DBM MRMC

For the stacked input format, the data must be organized in columns following the order: `ReaderID`, `ModalityID`, `CaseID`, `Truth` and `Rating`. The column names can be anything, but they must exist. The columns `ReaderID`, `ModalityID`, `CaseID` and `Rating` are defined as the same as JAFROC data file format. For the `Truth` column, use 0 for normal cases and 1 for abnormal. The delimiter in each line must be spaces or commas, but only one delimiter is allowed in a same data file. The data file can be saved as plain text file with *.txt or *.csv extensions.

iMRMC data format

The iMRMC data format contains two parts: the data description and ROC ratings.

1. Data description part includes free text information at the top of the file. Following information must be specified in this part as four lines

- `N0`, number of normal cases.
- `N1`, number of abnormal cases.
- `NR`, number of readers.
- `NM`, number of modalities.

The symbol and corresponding number are separated by colon (:). For example, 50 normal cases can be indicated by the line `N0 : 50`.

2. ROC ratings list starts at the line "`BEGIN DATA :`". The list has two parts: truth states and readers' ratings.

- Truth states has four columns in the following order:
 - `-1`, lines with `-1` in the first column indicate the lines are truth states lines.
 - `CaseID`, same as JAFROC data format.
 - `arbitrary string`, can be anything.
 - `Truth`, same as OR-DBM MRMC format.
- Readers' ratings list also contains four columns in the following order:
 - `ReaderID`, same as JAFROC data format. Note that `ReaderID` can be anything but "-1".
 - `CaseID`, same as JAFROC data format.
 - `ModalityID`, same as JAFROC data format.
 - `Ratings`, same as JAFROC data format.

Columns are separated by commas and the data are saved as plain text file with *.imrmc extension.

Note that JAFROC and OR-DBM MPMC data file have table headers, while the iMPMC data file does not.

Functions

CalculatePower	Calculate the statistical power with the given number of readers, number of cases and DBMP variances components.
CalculateSampleSize	Calculate required sample size for the specified dataset.
DBMAnalysis	Performs Dorfman-Berbaum-Metz (P _{db}) analysis for the specified dataset.
OperatingCharacteristics	Visualize the predicted ROC, AFROC, FROC and pdf (probability density function) curves for given search model parameters.
ORAnalysis	Performs Obuchowski-Rockette (P _{or}) analysis for the specified dataset.
OutputReport	Output the report of the analysis result.
PlotEmpiricalCurve	Plot empirical curves for specified modalities and readers in the dataset.
PlotParametricCurve	Plot parametric curves for specified modalities and readers in the dataset.
ReadDataFile	Read the dataset that will be analysis from data file
SampleSizeCalculator	Calculate required number of cases with the given number of readers and DBM variances components.
SaveDataFile	Save the dataset into a file as the specified format.

Author(s)

Xuetong Zhai, Dev Chakraborty

Maintainer: Xuetong Zhai <xuetong.zhai@gmail.com>