Package ‘mccf1’

November 11, 2019

Title Creates the MCC-F1 Curve and Calculates the MCC-F1 Metric and the Best Threshold

Version 1.1

Date 2019-11-11

Maintainer Chang Cao <kirin.cao@mail.utoronto.ca>

Depends R (>= 3.3.3), ggplot2

Imports ROCR

Description The MCC-F1 analysis is a method to evaluate the performance of binary classifications. The MCC-F1 curve is more reliable than the Receiver Operating Characteristic (ROC) curve and the Precision-Recall (PR) curve under imbalanced ground truth. The MCC-F1 analysis also provides the MCC-F1 metric that integrates classifier performance over varying thresholds, and the best threshold of binary classification.

License GPL (>= 2)

URL https://bitbucket.org/hoffmanlab/mccf1/

BugReports https://stackoverflow.com/questions/tagged/mccf1

RoxygenNote 6.0.1

NeedsCompilation no

Author Chang Cao [aut, cre], Michael Hoffman [aut], Davide Chicco [aut]

Repository CRAN

Date/Publication 2019-11-11 21:00:03 UTC

R topics documented:

autoplot.mccf1 .......................................................... 2
mccf1 ........................................................................ 2
summary.mccf1 ........................................................... 3

Index 5
### autoplot.mccf1

**Plot the MCC-F1 curve**

**Description**

`autoplot.mccf1()` plots the MCC-F1 curve using ggplot2.

**Usage**

```r
## S3 method for class 'mccf1'
autoplot(object, xlab = "F1 score", ylab = "normalized MCC", ...)
```

**Arguments**

- `object`: S3 object of class "mccf1" from the `mccf1()`
- `xlab`, `ylab`: x- and y- axis annotation (default: "F1 score","normalized MCC")
- `...`: further arguments passed to and from method `ggplot()`

**Value**

the ggplots object

**Examples**

```r
response <- c(rep(1, 1000), rep(0, 10000))
predictor <- c(rbeta(300, 12, 2), rbeta(700, 3, 4), rbeta(10000, 2, 3))
autoplot(mccf1(response, predictor))
```

### mccf1

**Perform MCCF1 analysis**

**Description**

`mccf1()` performs MCC (Matthews correlation coefficient)-F1 analysis for paired vectors of binary response classes and fractional prediction scores representing the performance of a binary classification task.

**Usage**

```r
mccf1(response, predictor)
```

**Arguments**

- `response`: numeric vector representing ground truth classes (0 or 1).
- `predictor`: numeric vector representing prediction scores (in the range [0,1]).
### summary.mccf1

**Summary**

S3 object of class "mccf1", a list with the following members: `thresholds`: vector of doubles describing the thresholds; `normalized_mcc`: vector of doubles representing normalized MCC for each threshold; `f1`: vector of doubles representing F1 for each threshold.

**Examples**

```r
response <- c(rep(1L, 1000L), rep(0L, 10000L))
set.seed(2017)
predictor <- c(rbeta(300L, 12, 2), rbeta(700L, 3, 4), rbeta(10000L, 2, 3))
x <- mccf1(response, predictor)
head(x$thresholds)
# [1] Inf 0.9935354 0.9931493 0.9930786 0.9925507 0.9900520
head(x$normalized_mcc)
# [1] NaN 0.5150763 0.5213220 0.5261152 0.5301566 0.5337177
head(x$f1)
# [1] NaN 0.001998002 0.003992016 0.005982054 0.007968127 0.009950249
```

---

### Description

`summary.mccf1()` calculates the MCC-F1 metric and the best threshold for a binary classification.

### Usage

```r
## S3 method for class 'mccf1'
summary(object, digits, bins = 100, ...)  
```

### Arguments

- **object**: S3 object of class "mccf1" object resulting from the function `mccf1()`
- **digits**: integer, used for number formatting with `signif`
- **bins**: integer, representing number of bins used to divide up the range of normalized MCC when calculating the MCC-F1 metric (default = 100L)
- **...**: other arguments ignored (for compatibility with generic)

### Value

data.frame that shows the MCC-F1 metric (in the range [0,1]) and the best threshold (in the range [0,1])
Examples

```r
response <- c(rep(1L, 1000L), rep(0L, 10000L))
set.seed(2017)
predictor <- c(rbeta(300L, 12, 2), rbeta(700L, 3, 4), rbeta(10000L, 2, 3))
## Not run: summary(mccf1(response, predictor))
# mccf1_metric best_threshold
# 0.3508904 0.786905
summary(mccf1(response, predictor), bins = 50)
# mccf1_metric best_threshold
# 0.3432971 0.786905
## Not run: summary(mccf1(response, predictor), digits = 3)
# mccf1_metric best_threshold
# 0.351 0.787
```
Index

autplot.mccf1, 2
mccf1, 2
signif, 3
summary.mccf1, 3