Package 'mri'

October 13, 2022

Title Modified Rand and Wallace Indices

Version 1.0.1

Description It provides functions to compute the values of different modifications of the Rand and Wallace indices. The indices are used to measure the stability or similarity of two partitions obtained on two different sets of units with a non-empty intercept. Splitting and merging of clusters can (depends on the selected index) have a different effect on the value of the indices. The indices are proposed in Cugmas and Ferligoj (2018) <http://ibmi.mf.uni-lj.si/mz/2018/no-1/Cugmas2018.pdf>.

Depends R (>= 3.1.0)

License GPL (>= 2)

LazyData true

RoxygenNote 6.1.0

NeedsCompilation no

Author Marjan Cugmas [aut, cre]

Maintainer Marjan Cugmas <marjan.cugmas@fdv.uni-lj.si>

Repository CRAN

Date/Publication 2019-03-07 12:50:12 UTC

R topics documented:

---

examples   *Examples of Misclassifications of Units*

Description

Some examples of misclassifications of units in clusters of the first and second partitions. The data are the same as used in Cugmas and Ferligoj (2018) (see Figure 4 in the paper) (the exception are examples d, h, l and p which are random and therefore they can slightly differ from those in the paper).

Usage

```
data("examples")
```

Format

The data are in a list format. Each element of the list is a contingency table. The newcomers and outgoers are not denoted.

Details

Element of the list (example in Figure 4 and in Table 3 in Cugmas and Ferligoj (2018))

1 (A), 2 (B), 3(C), 4 (D)

13 (E), 14 (F), 15 (G), 16 (H)

5 (I), 6 (J), 7 (K), 8 (L)

9 (M), 10 (N), 11 (0), 12 (P)

Source

Cugmas, M., & Ferligoj, A. (2018). Comparing two partitions of non-equal sets of units. Advances in Methodology and Statistics, 15(1), 1-21.

---

fromTableToVectors   *From Contingency Table to Data Frame*

Description

It covert a given contingency table to a data frame.

Usage

```
fromTableToVectors(cont.table)
```

Arguments

cont.table	contingency table (a data frame with rownames and columnnames)

Value

A data frame with n rows and 2 columns. The first column corresponds to the rows of the contingency table while the second column corresponds to the columns of the contingency table.

Author(s)

Marjan Cugmas

Examples

```
data <- rbind(c(0, 10, 0, 0, 0),
                c(0, 10, 0, 0, 0),
                c(0, 0, 10, 0, 0),
                c(0, 0,  0, 5, 5))
rownames(data) <- 1:4
colnames(data) <- 1:5
fromTableToVectors(cont.table = data)
```

MRI   *Modified Adjusted Rand Index*

Description

They are used to compute the value of the Modified Rand Index and the Modified Adjusted Rand Index. They consider two partitions which are usually obtained on two sets of units where the intercept is non-empty or where one set of units is a subset of another set of units.

Because two vectors U and V (partitions) do not have the same length, the cluster of units, which are not present in the partition V (outgoers), need to be added to the partition V (denoted as V') and/or the cluster of units, which are not present in the partition U (newcomers), need to be added to the partition U (denoted as U'). The added values have to be NA.

Usage

```
MRI(U, V)
MARI(U, V, k)
```

Arguments

U	partition U or U'
V	partition V or V'
k	the number of iterations to estimate the expected value of the index in the case of two random and independent partitions

Value

The functions return the value of the Modified Rand Index or the value of the Modified Adjusted Rand Index. The expected value of the (Modified) Adjusted Rand Index is 0 in the case of two random and independent partitions. The maximum value of the index is 1. Higher value indicates more similar (stable) partitions. Both splitting of clusters and merging of clusters lower the value of the indices.

Note

The special cases of the modified indices (when only outgoers or only newcomers are present) are automatically considered within these functions.

Author(s)

Marjan Cugmas

References

Cugmas, M., & Ferligoj, A. (2018). Comparing two partitions of non-equal sets of units. Advances in Methodology and Statistics, 15(1), 1-21.

Examples

```
# Examples from Cugmas and Ferligoj (2018) paper:
data(examples)
# increase k in real analyses
# EXAMPLES: A, B, C, D
par(mfrow = c(4, 4))
for (i in 1:4){
  U <- fromTableToVectors(examples[[i]])[,1]
  V <- fromTableToVectors(examples[[i]])[,2]

  cat("MARI", MARI(U = U, V = V, k = 100), "\n")
}

# EXAMPLES: E, F, G, H
for (i in 13:16){
  U <- fromTableToVectors(examples[[i]])[,1]
  V <- fromTableToVectors(examples[[i]])[,2]

  U[which(max(as.numeric(as.character(U))) == as.numeric(as.character(U)))] <- NA
  V[which(max(as.numeric(as.character(V))) == as.numeric(as.character(V)))] <- NA

  cat("MARI", MARI(U = U, V = V, k = 100), "\n")
}

# EXAMPLES: I, J, K, L
for (i in 5:8){
  U <- fromTableToVectors(examples[[i]])[,1]
  V <- fromTableToVectors(examples[[i]])[,2]
```

```
    V[which(max(as.numeric(as.character(V))) == as.numeric(as.character(V)))] <- NA

    cat("MARI", MARI(U = U, V = V, k = 100), "\n")
}

# EXAMPLES: M, N, O, P
for (i in 9:12){
  U <- fromTableToVectors(examples[[i]])[,1]
  V <- fromTableToVectors(examples[[i]])[,2]

  U[which(max(as.numeric(as.character(U))) == as.numeric(as.character(U)))] <- NA

  cat("MARI", MARI(U = U, V = V, k = 100), "\n")
}
```

MW1   *Modified Adjusted Wallace Index 1*

Description

The functions are used to compute the value of the Modified Wallace Index 1 and the Modified
Adjusted Wallace Index 1. They consider two partitions which are usually obtained on two sets of
units where the intercept is non-empty or where one set of units is a subset of another set of units.

Because two vectors U and V (partitions) do not have the same length, the cluster of units, which
are not present in the partition V (outgoers), need to be added to the partition V (denoted as V')
and/or the cluster of units, which are not present in the partition U (newcomers), need to be added
to the partition U (denoted as U'). The added values have to be NA.

Usage

```
MW1(U, V)
MAW1(U, V, k)
```

Arguments

U	partition U or U'
V	partition V or V'
k	the number of iterations to estimate the expected value of the index in the case of two random and independent partitions

Value

The functions return the value of the Modified Wallace Index 1 or the value of the Modified Adjusted
Wallace Index 1. The expected value of the (Modified) Adjusted Wallace indices 1 is 0 in the case
of two random and independent partitions. The maximum value of the index is 1. Higher value
indicates more similar (stable) partitions. Splitting of clusters lower the value of the indices while
merging does not.

Note

The special cases of the modified indices (when only outgoers or only newcomers are present) are automatically considered within these functions.

Author(s)

Marjan Cugmas

References

Cugmas, M., & Ferligoj, A. (2018). Comparing two partitions of non-equal sets of units. Advances in Methodology and Statistics, 15(1), 1-21.

Examples

```
# Examples from Cugmas and Ferligoj (2018) paper:
data(examples)
# EXAMPLES: A, B, C, D
par(mfrow = c(4, 4))
for (i in 1:4){
  U <- fromTableToVectors(examples[[i]])[,1]
  V <- fromTableToVectors(examples[[i]])[,2]

  cat("MAWI1", MAW1(U = U, V = V), "\n")
}

# EXAMPLES: E, F, G, H
for (i in 13:16){
  U <- fromTableToVectors(examples[[i]])[,1]
  V <- fromTableToVectors(examples[[i]])[,2]

  U[which(max(as.numeric(as.character(U))) == as.numeric(as.character(U)))] <- NA
  V[which(max(as.numeric(as.character(V))) == as.numeric(as.character(V)))] <- NA

  cat("MAWI1", MAW1(U = U, V = V), "\n")
}

# EXAMPLES: I, J, K, L
for (i in 5:8){
  U <- fromTableToVectors(examples[[i]])[,1]
  V <- fromTableToVectors(examples[[i]])[,2]

  V[which(max(as.numeric(as.character(V))) == as.numeric(as.character(V)))] <- NA

  cat("MAWI1", MAW1(U = U, V = V), "\n")
}

# EXAMPLES: M, N, O, P
for (i in 9:12){
  U <- fromTableToVectors(examples[[i]])[,1]
  V <- fromTableToVectors(examples[[i]])[,2]
```

```
  U[which(max(as.numeric(as.character(U))) == as.numeric(as.character(U)))] <- NA

  cat("MAWI1", MAW1(U = U, V = V), "\n")
}
```

MW2   *Modified Adjusted Wallace Index 2*

Description

The functions are used to compute the value of the Modified Wallace Index 2 and the Modified Adjusted Wallace Index 2. They consider two partitions which are usually obtained on two sets of units where the intercept is non-empty or where one set of units is a subset of another set of units.

Because two vectors U and V (partitions) do not have the same length, the cluster of units, which are not present in the partition V (outgoers), need to be added to the partition V (denoted as V') and/or the cluster of units, which are not present in the partition U (newcomers), need to be added to the partition U (denoted as U'). The added values have to be NA.

Usage

```
MW2(U, V)
MAW2(U, V, k)
```

Arguments

U	partition U or U'
V	partition V or V'
k	the number of iterations to estimate the expected value of the index in the case of two random and independent partitions

Value

The functions return the value of the Modified Wallace Index 2 or the value of the Modified Adjusted Wallace Index 2. The expected value of the (Modified) Adjusted Wallace indices 2 is 0 in the case of two random and independent partitions. The maximum value of the index is 2. Higher value indicates more similar (stable) partitions. Merging of clusters lowers the value of the indices while splitting does not.

Note

The special cases of modified indices (when only outgoers or only newcomers are present) are automatically considered within these functions.

Author(s)

Marjan Cugmas

References

Cugmas, M., & Ferligoj, A. (2018). Comparing two partitions of non-equal sets of units. Advances in Methodology and Statistics, 15(1), 1-21.

Examples

```
# Examples from Cugmas and Ferligoj (2018) paper:
data(examples)
# EXAMPLES: A, B, C, D
par(mfrow = c(4, 4))
for (i in 1:4){
  U <- fromTableToVectors(examples[[i]])[,1]
  V <- fromTableToVectors(examples[[i]])[,2]

  cat("MAWI2", MAW2(U = U, V = V), "\n")
}

# EXAMPLES: E, F, G, H
for (i in 13:16){
  U <- fromTableToVectors(examples[[i]])[,1]
  V <- fromTableToVectors(examples[[i]])[,2]

  U[which(max(as.numeric(as.character(U))) == as.numeric(as.character(U)))] <- NA
  V[which(max(as.numeric(as.character(V))) == as.numeric(as.character(V)))] <- NA

  cat("MAWI2", MAW2(U = U, V = V), "\n")
}

# EXAMPLES: I, J, K, L
for (i in 5:8){
  U <- fromTableToVectors(examples[[i]])[,1]
  V <- fromTableToVectors(examples[[i]])[,2]

  V[which(max(as.numeric(as.character(V))) == as.numeric(as.character(V)))] <- NA

  cat("MAWI2", MAW2(U = U, V = V), "\n")
}

# EXAMPLES: M, N, O, P
for (i in 9:12){
  U <- fromTableToVectors(examples[[i]])[,1]
  V <- fromTableToVectors(examples[[i]])[,2]

  U[which(max(as.numeric(as.character(U))) == as.numeric(as.character(U)))] <- NA

  cat("MAWI2", MAW2(U = U, V = V), "\n")
}
```

Index