Package 'InterVA4'

November 13, 2019

Type Package

Title Replicate and Analyse 'InterVA4'

Version 1.7.6

Date 2019-11-11

Author Zehang Richard Li, Tyler McCormick, Sam Clark, Peter Byass

Maintainer Zehang Richard Li <lizehang@gmail.com>

Description Provides an R version of the 'InterVA4' software (<http://www.interva.net>) for coding cause of death from verbal autopsies. It also provides simple graphical representation of individual and population level statistics.

License GPL-3

RoxygenNote 6.1.1

Suggests testthat

NeedsCompilation no

Repository CRAN

Date/Publication 2019-11-13 05:40:08 UTC

R topics documented:

InterVA4-package	*Perform InterVA4 algorithm and provide graphical summarization of COD distribution.*

Description

Computes individual cause of death and population cause-specific mortality fractions using the InterVA4 algorithm. Provides a simple graphical representation of the result.

Details

To get the most up-to-date version of the package, as well as the past versions, please check the github repository at: https://github.com/richardli/InterVA4/

Package:	InterVA4
Type:	Package
Version:	1.6
Date:	2015-08-29
License:	GPL-2

Author(s)

Zehang Li, Tyler McCormick, Sam Clark

Maintainer: Zehang Li <lizehang@uw.edu>

References

http://www.interva.net/

Examples

```
data(SampleInput)
sample.output <- InterVA(SampleInput, HIV = "h", Malaria = "v", write=FALSE)
```

causetext	*Translation list of COD codes*

Description

This is the translation of COD abbreviation codes into their corresponding full names.

Format

A data frame with the translation of COD codes to their names on 68 CODs (both the version of COD only and COD with group code).

Examples

```
data(causetext)
```

CSMF   *Summarize and plot a population level distribution of va probabilities.*

Description

The function takes input of a list of va object and produces a summary plot for the population distribution.

Usage

```
CSMF(va, top.aggregate = NULL, InterVA.rule = FALSE, noplot = FALSE,
  type = "bar", top.plot = 10, min.prob = 0, ...)
```

Arguments

va	The list of va object to summarize.
top.aggregate	Integer indicating how many causes from the top need to go into summary. The rest of the probabilities goes into an extra category "Undetermined". When set to NULL, default is all causes to be considered. This is only used when InterVA.rule set to "FALSE".
InterVA.rule	If it is set to "TRUE", only the top 3 causes reported by InterVA4 is calculated into CSMF as in InterVA4. The rest of probabilities goes into an extra category "Undetermined". Default set to "FALSE".
noplot	A logical value indicating whether the plot will be shown. If it is set to "TRUE", only the CSMF will be returned.
type	An indicator of the type of chart to plot. "pie" for pie chart; "bar" for bar chart.
top.plot	the maximum number of causes to plot in bar plot
min.prob	The minimum probability that is to be plotted in bar chart, or to be labeled in pie chart.
...	Arguments to be passed to/from graphic function [barplot](), [pie](), and more graphical paramters (see [par]()). They will affect the main title, size and font of labels, and the radius of the pie chart.

Value

dist.cod	The population probability of CODs.

Author(s)

Zehang LI, Tyler McCormick, Sam Clark

See Also

CSMF.interVA4

Examples

```
data(SampleInput)
sample.output <- InterVA(SampleInput, HIV = "h", Malaria = "v", write=FALSE)

## Get CSMF without plots
population.summary <- CSMF(sample.output$VA, noplot = TRUE)


## Get CSMF by considering only top 3 causes for each death.
population.summary <- CSMF(sample.output$VA, top.aggregate = 3, noplot = TRUE)

## Get CSMF by considering only top 3 causes reported by InterVA.
## This is equivalent to using CSMF.interVA4() command Note that
## it's different from using all top 3 causses, since they may not
## all be reported
CSMF.summary <- CSMF(sample.output, InterVA.rule = TRUE,
   noplot = TRUE)

## Population level summary using pie chart
CSMF.summary2 <- CSMF(sample.output, type = "pie",
 min.prob = 0.01, main = "population COD distribution using pie chart",
 clockwise = FALSE, radius = 0.7, cex = 0.7, cex.main = 0.8)

## Population level summary using bar chart
CSMF.summary3 <- CSMF(sample.output, type = "bar",
  min.prob = 0.01, main = "population COD distribution using bar chart",
  cex.main = 1)
CSMF.summary4 <- CSMF(sample.output, type = "bar",
  top.plot = 5, main = "Top 5 population COD distribution",
  cex.main = 1)
```

CSMF.interVA4	*Summarize population level cause-specific mortality fraction as In-terVA4 suggested.*

Description

The function takes input of a list of va object and calculates the cause-specific mortality fraction. It only calculates CSMF as aggregation of up to the third largest causes.

Usage

```
CSMF.interVA4(va)
```

Arguments

va   The list of va object to summarize.

Value

dist.cod   The cause-specific mortality fraction (including undetermined category).

Author(s)

Zehang LI, Tyler McCormick, Sam Clark

See Also

[CSMF](#)

Examples

```
data(SampleInput)
sample.output <- InterVA(SampleInput, HIV = "h", Malaria = "v", write=FALSE)
## Get CSMF without plots
csmf<- CSMF.interVA4(sample.output$VA)
data(SampleInput)
```

InterVA   *Provide InterVA4 analysis on the data input.*

Description

This function implements the algorithm in the InterVA4 software. It produces individual cause of death and population cause-specific mortality fractions.

Usage

```
InterVA(Input, HIV, Malaria, directory = NULL, filename = "VA_result",
  output = "classic", append = FALSE, groupcode = FALSE,
  replicate = FALSE, replicate.bug1 = FALSE, replicate.bug2 = FALSE,
  write = TRUE, ...)
```

Arguments

Input	A matrix input, or data read from csv files in the same format as required by InterVA4. Sample input is included as data(SampleInput).
HIV	An indicator of the level of prevalence of HIV. The input should be one of the following: "h"(high),"l"(low), or "v"(very low).
Malaria	An indicator of the level of prevalence of Malaria. The input should be one of the following: "h"(high),"l"(low), or "v"(very low).
directory	The directory to store the output from InterVA4. It should either be an existing valid directory, or a new folder to be created. If no path is given, the current working directory will be used.
filename	The filename the user wish to save the output. No extension needed. The output is in .csv format by default.
output	"classic": The same deliminated output format as InterVA4; or "extended": deliminated output followed by full distribution of cause of death proability.
append	A logical value indicating whether or not the new output should be appended to the existing file.
groupcode	A logical value indicating whether or not the group code will be included in the output causes.
replicate	A logical value indicating whether or not the calculation should replicate original InterVA4 software (version 4.02) exactly. If replicate = F, causes with small probability are not dropped out of calculation in intermediate steps, and a possible bug in original InterVA4 implementation is fixed. If replicate=T, then the output values will be exactly as they would be from calling the InterVA4 program (version 4.02). If replicate=F, the output values will be the same as calling the InterVA4 program (version 4.03). Since version 1.7.3, setting replicate to be FALSE also includes changes to data checking rules and pre-set conditional probabilities to be the same as the official version 4.03 software. Since version 1.6, two control variables are added to control the two bugs respectively. Setting this to TRUE will overwrite both to TRUE.
replicate.bug1	This logical indicator controls whether or not the bug in InterVA4.2 involving the symptom "skin_les" will be replicated or not. It is suggested to set to FALSE.
replicate.bug2	This logical indicator controls whether the causes with small probability are dropped out of calculation in intermediate steps or not. It is suggested to set to FALSE.
write	A logical value indicating whether or not the output (including errors and warnings) will be saved to file.
...	not used

Details

InterVA performs the same tasks as the InterVA4. The output is saved in a .csv file specified by user. The calculation is based on the conditional and prior distribution of 68 CODs. The function also could save the full probability distibution of each individual to file. All information about each individual is saved to a va class object.

Be careful if the input file does not match InterVA input format strictly. The function will run normally as long as the number of symptoms are correct. Any inconsistent symptom names will be printed in console as warning. If there's wrong match of symptom from warning, please change in the input to correct orders.

Value

ID	identifier from batch (input) file
MALPREV	selected malaria prevalence
HIVPREV	selected HIV prevalence
PREGSTAT	most likely pregnancy status
PREGLIK	likelihood of PREGSTAT
PRMAT	likelihood of maternal death
INDET	indeterminate outcome
CAUSE1	most likely cause
LIK1	likelihood of 1st cause
CAUSE2	second likely cause
LIK2	likelihood of 2nd cause
CAUSE3	third likely cause
LIK3	likelihood of 3rd cause
wholeprob	full distribution of causes of death

Author(s)

Zehang Li, Tyler McCormick, Sam Clark

References

http://www.interva.net/

See Also

[InterVA.plot](#)

Examples

```
data(SampleInput)
## to get easy-to-read version of causes of death make sure the column
## orders match interVA4 standard input this can be monitored by checking
## the warnings of column names

sample.output1 <- InterVA(SampleInput, HIV = "h", Malaria = "l", write=FALSE, replicate = FALSE)

## to get causes of death with group code for further usage
sample.output2 <- InterVA(SampleInput, HIV = "h", Malaria = "l", write=FALSE,
    replicate = FALSE, groupcode = TRUE)
```

InterVA.plot   *Plot a individual level distribution of va probabilities.*

Description

The function takes input of a single va object and produces a summary plot for it.

Usage

```
InterVA.plot(va, type = "bar", min.prob = 0.01, ...)
```

Arguments

va	A va object
type	An indicator of the type of chart to plot. "pie" for pie chart; "bar" for bar chart.
min.prob	The minimum probability that is to be plotted in bar chart, or to be labeled in pie chart.
...	Arguments to be passed to/from graphic function barplot, pie, and more graphical paramters (see par). They will affect the main title, size and font of labels, and the radius of the pie chart.

See Also

CSMF

Examples

```
data(SampleInput)
sample.output <- InterVA(SampleInput, HIV = "h", Malaria = "v", write=FALSE)

## Individual level summary using pie chart
InterVA.plot(sample.output$VA[[7]], type = "pie", min.prob = 0.01,
    main = "1st sample VA analysis using pie chart", clockwise = FALSE,
    radius = 0.6, cex = 0.6, cex.main = 0.8)


## Individual level summary using bar chart
InterVA.plot(sample.output$VA[[7]], type = "bar", min.prob = 0.01,
    main = "2nd sample VA analysis using bar chart", cex.main = 0.8)
```

Population.summary   *Summarize and plot a population level distribution of va probabilities.*

Description

This function has been deprecated as of version 1.6. Use 'CSMF' instead.

Usage

```
Population.summary(va, top.aggregate = NULL, InterVA.rule = FALSE,
  noplot = FALSE, type = "bar", min.prob = 0.01, ...)
```

Arguments

va	The list of va object to summarize.
top.aggregate	Integer indicating how many causes from the top need to go into summary. The rest of the probabilities goes into an extra category "Undetermined". When set to NULL, default is all causes to be considered. This is only used when InterVA.rule set to "FALSE".
InterVA.rule	If it is set to "TRUE", only the top 3 causes reported by InterVA4 is calculated into CSMF as in InterVA4. The rest of probabilities goes into an extra category "Undetermined". Default set to "FALSE".
noplot	A logical value indicating whether the plot will be shown. If it is set to "TRUE", only the CSMF will be returned.
type	An indicator of the type of chart to plot. "pie" for pie chart; "bar" for bar chart.
min.prob	The minimum probability that is to be plotted in bar chart, or to be labeled in pie chart.
...	Arguments to be passed to/from graphic function barplot, pie, and more graphical paramters (see par). They will affect the main title, size and font of labels, and the radius of the pie chart.

Value

dist.cod	The population probability of CODs.

Author(s)

Zehang LI, Tyler McCormick, Sam Clark

See Also

CSMF.interVA4

---

print.interVA_summary *Print method for summary of the results obtained from InterVA4 algorithm*

Description

This function prints the summary message of the fitted results.

Usage

```
## S3 method for class 'interVA_summary'
print(x, ...)
```

Arguments

x   summary of InterVA results

...   not used

---

probbase   *Conditional probability of InterVA4.02*

Description

This is the table of conditional probabilities of symptoms given CODs. The values are from InterVA-4.02.

Format

A data frame with 246 observations on 81 variables. Each observation is the conditional probability.

Examples

```
data(probbase)
```

---

probbase3	*Conditional probability of InterVA4.03*

Description

This is the table of conditional probabilities of symptoms given CODs. The values are from InterVA-4.03.

Format

A data frame with 246 observations on 81 variables. Each observation is the conditional probability.

Examples

```
data(probbase)
```

SampleInput	*10 records of Sample Input*

Description

This is a dataset consisting of 10 arbitrary sample input deaths in the acceptable format of InterVA4. Any data that needs to be analyzed by this package should be in the same format. The orders of the input fields must not be changed.

Format

10 arbitrary input records.

Examples

```
data(SampleInput)
```

summary.interVA	*Summary of the results obtained from InterVA4 algorithm*

Description

This function prints the summary message of the fitted results.

Usage

```
## S3 method for class 'interVA'
summary(object, top = 5, id = NULL,
  InterVA.rule = TRUE, ...)
```

Arguments

object	fitted object from InterVA()
top	number of top CSMF to show
id	the ID of a specific death to show
InterVA.rule	If it is set to "TRUE", only the top 3 causes reported by InterVA4 is calculated into CSMF as in InterVA4. The rest of probabilities goes into an extra category "Undetermined". Default set to "TRUE".
...	not used

References

http://www.interva.net/

Examples

```
data(SampleInput)
## to get easy-to-read version of causes of death make sure the column
## orders match interVA4 standard input this can be monitored by checking
## the warnings of column names

sample.output1 <- InterVA(SampleInput, HIV = "h", Malaria = "l", write=FALSE, replicate = FALSE)

summary(sample.output1)
summary(sample.output1, top = 10)
summary(sample.output1, id = "100532")
```

Index