Package ‘deductive’

April 10, 2019

Maintainer Mark van der Loo <mark.vanderloo@gmail.com>
License GPL-3
Title Data Correction and Imputation Using Deductive Methods
LazyData no
Type Package
LazyLoad yes
Description Attempt to repair inconsistencies and missing values in data records by using information from valid values and validation rules restricting the data.
Version 0.1.3
Depends R (>= 3.2.0)
URL https://github.com/data-cleaning/deductive
BugReports https://github.com/data-cleaning/deductive/issues
Imports methods, lintools (>= 0.1.2), validate, stringdist
Suggests testthat, knitr
RoxygenNote 6.1.1
NeedsCompilation yes
Author Mark van der Loo [cre, aut],
Edwin de Jonge [aut]
Repository CRAN
Date/Publication 2019-04-10 13:55:46 UTC

R topics documented:
correct_typos .................................................. 2
deductive ....................................................... 3
impute_lr ......................................................... 3

Index 5
Correct typos in restricted numeric data

Description

Attempt to fix violations of linear (in)equality restrictions imposed on a record by replacing values with values that differ from the original values by typographical errors.

Usage

```r
correct_typos(dat, x, ...)
```

```r
## S4 method for signature 'data.frame,validator'
correct_typos(dat, x, fixate = NULL,
    eps = 1e-08, maxdist = 1, ...)
```

Arguments

- `dat`  
  An R object holding numeric (integer) data.
- `x`  
  An R object holding linear data validation rules
- `...`  
  Options to be passed to `stringdist` which is used to determine the typographic distance between the original value and candidate solutions. By default, the optimal string alignment distance is used, with all weights equal to one.
- `fixate`  
  [character] vector of variable names that may not be changed
- `eps`  
  [numeric] maximum roundoff error
- `maxdist`  
  [numeric] maximum allowed typographical distance

Value

`dat`, with values corrected.

Details

The algorithm works by proposing candidate replacement values and checking whether they are likely to be the result of a typographical error. A value is accepted as a solution when it resolves at least one equality violation. An equality restriction `a.x=b` is considered satisfied when `abs(a.x-b)<eps`. Setting `eps` to one or two units of measurement allows for robust typographical error detection in the presence of roundoff-errors.

The algorithm is meant to be used on numeric data representing integers.

References

- The first version of the algorithm was described by S. Scholtus (2009). Automatic correction of simple typing errors in numerical data with balance edits. Statistics Netherlands, Discussion Paper 09046
The generalized version of this algorithm that is implemented for this package is described in M. van der Loo, E. de Jonge and S. Scholtus (2011). Correction of rounding, typing and sign errors with the deducorrect package. Statistics Netherlands, Discussion Paper 2011019

Examples

```r
library(validate)

# example from section 4 in Scholtus (2009)

v <- validate::validator(
x1 + x2 == x3,
x2 == x4,
x5 + x6 + x7 == x8,
x3 + x8 == x9,
x9 - x10 == x11
)

dat <- read.csv(textConnection("x1, x2, x3, x4, x5, x6, x7, x8, x9, x10, x11
1452, 116, 1568, 116, 323, 76, 12, 411, 1979, 1842, 137
1452, 116, 1568, 161, 323, 76, 12, 411, 1979, 1842, 137
1452, 116, 1568, 161, 323, 76, 12, 411, 19979, 1842, 137
1452, 116, 1568, 161, 0, 0, 0, 411, 19979, 1842, 137
1452, 116, 1568, 161, 323, 76, 12, 0, 19979, 1842, 137")
))
cor <- correct_typos(dat, v)
dat - cor
```

---

deductive  

**Deductive Data Correction and Imputation**

**Description**

Deductive Data Correction and Imputation

**impute_lr**

*Impute values derived from linear (in)equality restrictions.*
**Description**

Partially filled records \( x \) under (in)equality restrictions may reveal unique imputation solutions when the system of linear inequalities is reduced by substituting observed values. This function applies a number of fast heuristic methods before deriving all variable ranges and unique values.

**Usage**

\[
\text{impute} \_\text{lr} (\text{dat}, x, \ldots)
\]

```r
## S4 method for signature 'data.frame,validator'
impute\_lr(dat, x, \ldots)
```

**Arguments**

- `dat`: an R object carrying data
- `x`: an R object carrying validation rules
- `\ldots`: arguments to be passed to other methods.

**Examples**

```r
v <- validate::validator(y >= 2, y + z >= 3, x + y <= 0)
dat <- data.frame(x = NA_real_, y = NA_real_, z = NA_real_)
impute\_lr(dat, v)
```
Index

correct_typos, 2
correct_typos, data.frame, validator-method
  (correct_typos), 2

deductive, 3
deductive-package (deductive), 3

impute_lr, 3
impute_lr, data.frame, validator-method
  (impute_lr), 3

stringdist, 2