Package ‘ccrs’

October 12, 2022

Type Package
Title Correct and Cluster Response Style Biased Data
Version 0.1.0
Author Mariko Takagishi [aut, cre]
Maintainer Mariko Takagishi <m.takagishi0728@gmail.com>
Description Functions for performing Correcting and Clustering response-style-bias-
    preference data (CCRS). The main functions are correct.RS() for correcting for re-
    sponse styles, and ccrs() for simultaneously correcting and content-based clustering. The proce-
    dure begin with making rank-ordered boundary data from the given preference matrix us-
    ing a function called create.ccrsdata(). Then in correct.RS(), the response style is cor-
    rected as follows: the rank-ordered boundary data are smoothed by I-spline func-
    tions, the given preference data are transformed by the smoothed functions. The result-
    ing data matrix, which is considered as bias-corrected data, can be used for any data analy-
    sis methods. If one wants to cluster respondents based on their indicated preferences (content-
    based clustering), ccrs() can be applied to the given (response-style-biased) prefer-
    ence data, which simultaneously corrects for response styles and clusters respon-
    dents based on the contents. Also, the correction result can be checked by plot.crs() function.
License GPL (>= 2)
Depends R (>= 3.5.0)
Imports cds, colorspace, dplyr, graphics, limSolve, lsbclust, methods,
    msm, parallel, stats, utils
NeedsCompilation yes
Encoding UTF-8
RoxygenNote 6.1.1
Repository CRAN
Date/Publication 2019-03-04 17:10:02 UTC

R topics documented:

ccrs-package ................................................................. 2
ccrs ................................................................. 2
correcting and clustering preference data in the presence of response style bias.

Description
Corrects and clusters response-style-biased data.

Author(s)
Mariko Takagishi

References

ccrs
Correcting and Clustering response style biased data

Description
Applies CCRS to ccrsdata.list.

Usage
cccrs(ccrsdata.list,K=K,lam=lam, tandem.initial=FALSE,
       tol = 1e-5, maxit = 50, trace = 1, nstart = 3, parallel=F,verbose=T)

Arguments
ccrsdata.list A list generated by create.ccrsdata.
K An integer indicating the number of content-based clusters used for CCRS estimation.
lam A numeric value indicating lambda used for CCRS estimation.
tandem.initial A logical value indicating whether the 1st initial value is generated by CCRS tandem initialization. See Section 3.3 in the paper for the detail.
tol  A numeric value indicating the absolute convergence tolerance
maxit  An integer indicating the maximum number of iterations
trace  An non-negative integer. If positive, tracing information on the progress of the optimization is produced. Higher values produce more tracing information.
nstart  An integer indicating the number of random initial values.
parallel  A logical value indicating parallellization over starts is used.
verbose  A logical value indicating if the progress is printed during the iteration (only when parallel==FALSE).

Value

Returns a list with the following elements.

G  A K by m matrix of content-based cluster centroid.
cls.cont.vec  A vector of integers (from 1:K) indicating the content-based cluster to which each respondent is allocated.
opt.obval  An optimal value of objective function.
crs.list  A list of class crs, same as the one generated by correct.rs.

References


See Also

correct.rs

Examples

###data setting
n <- 30 ; m <- 10 ; H.true <- 2 ; K.true <- 2 ; q <- 5
datagene <- generate.rsdata(n=n,m=m,K.true=K.true,H.true=H.true,q=q,clustered.rs = TRUE)
###obtain n x m data matrix
X <- datagene$X
ccrsdata.list <- create.ccrsdata(X=q=q)
###CCRS
lam <- 0.8 ; K <- 2
ccrs.list <- ccrs(ccrsdata.list,K=K,lam=lam)
###check content-based clustering result
ccrs.list$cls.cont.vec
###check correction result
plot(ccrs.list$crs.list)
**convert.X2F**  
*Convert data matrix to rank-ordered boundary data*

**Description**
Converts data matrix to rank-ordered boundary data.

**Usage**
```r
correct.rs( X, q=q )
```

**Arguments**
- `X`: An n by m categorical data matrix.
- `q`: An integer indicating the maximum rating.

**Value**
An n by q-1 scaled rank-ordered boundary data.

**correct.rs**  
*Correct response-style-biased data*

**Description**
Corrects response-style-biased data, given `ccrsdata.list` created by `create.ccrsdata`.

**Usage**
```r
correct.rs( ccrsdata.list )
```

**Arguments**
- `ccrsdata.list`: A list generated by `create.ccrsdata`, which contains `Fmat`, `Mmat.q1`, `Mmat.q` and `X`.

**Value**
Returns an object of `crs` with the following elements.
- `Beta`: An n by q-1 matrix of coefficients for response functions.
- `Y.hat`: An n by m matrix of corrected data matrix.
- `MB`: An n by q matrix of values of response functions evaluated at the midpoint between boundaries.
create.ccrsdata

References


See Also

create.ccrsdata

Examples

### data setting
n <- 30 ; m <- 10 ; H.true <- 2 ; K.true <- 2 ; q <- 5
datagene <- generate.rsdata(n=n,m=m,K.true=K.true,H.true=H.true,q=q,clustered.rs = TRUE)
### obtain n x m data matrix
X <- datagene$X
crsdata.list <- create.ccrsdata(X,q=q)
crs.list <- correct.rs(crsdata.list)

create.ccrsdata Create a dataset for CCRS

Description

Creates a dataset for CCRS from a preference data matrix.

Usage

create.ccrsdata(X,q=q)

Arguments

X An n by m categorical data matrix.
q An integer indicating the maximum rating.

Details

For the difference between Mmat.q and Mmat.q1 in the resulting list, see Section 3.2 in reference paper.

Value

Returns a list with the following elements.

Fmat An n by q-1 matrix of scaled rank-ordered boundary data.
Mmat.q1 A q-1 by 3+1 matrix of I-spline basis functions, evaluated at the boundaries. +1 indicates all 0 intercepts.
Mmat.q A q by 3+1 matrix of I-spline basis functions, evaluated at the midpoints between boundaries.
X An n by m categorical data matrix same as the input X.
**generate.rsdata**

Simulate preference data to apply CCRS

**Description**

Simulates artificial preference data containing content-based (and response-style-based) clusters.

**Usage**

```r
generate.rsdata(n=n,m=m,q=q,K.true=K.true,H.true=NULL,clustered.rs=FALSE, 
cls.cont.vec=NULL,cls.rs.vec=NULL,savedata=FALSE)
```

**Arguments**

- `n` An integer indicating the number of respondents.
- `m` An integer indicating the number of items.
- `q` An integer indicating the maximum rating.
- `K.true` An integer indicating the true number of content-based clusters for `n` respondents.
- `H.true` An integer indicating the true number of response-style-based clusters for `n` respondents. This is needed when `clustered.rs=TRUE`.
- `clustered.rs` A logical value indicating whether response-style-based cluster structure exists in generated data. If TRUE, coefficients of L-spline are generated by response-style-based clusters. The default is `clustered.rs=FALSE`.
- `cls.cont.vec` A vector of integers (from 1:K.true) of length n indicating the content-based cluster to which each respondent is allocated in artificial data. If it’s NULL, it is generated automatically.
- `cls.rs.vec` A vector of integers (from 1:H.true) of length n indicating the response-style-based clusters. If it’s NULL and `clustered.rs==T`, it is generated randomly.
- `savedata` A logical value indicating whether artificial data are saved as csv files. The default is `savedata=FALSE`.

**References**


**See Also**

`correct.rs`
Value

A list with the following elements:

- **X**: An n by m matrix of categorical variables.
- **X.star**: An n by m matrix of true preference data \(X^*\).
- **X.nors**: An n by m matrix of categorical variables transformed by reference boundaries.
- **cls.cont.vec**: A vector of integers (from 1:H.true) indicating content-based clusters used to generate artificial data.
- **cls.rs.vec**: A vector of integers (from 1:H.true) indicating response-style-based clusters used to generate artificial data.

References


See Also

create.ccrsdata

Examples

```r
# data setting
n <- 30 ; m <- 10 ; H.true <- 2 ; K.true <- 2 ; q <- 5
datagene <- generate.rsdata(n=n,m=m,K.true=K.true,H.true=H.true,q=q,clustered.rs = TRUE)
# obtain n x m data matrix
X <- datagene$X
```

---

**plot.crs**

*Plot crs objects*

**Description**

Plots results of correction (1st plot: estimated response functions, 2nd plot: coefficient plot. See Appendix A of the reference paper for the 2nd plot).

**Usage**

```r
## S3 method for class 'crs'
plot(x, H = NULL, cls.rs.vec = NULL, ...)
```
Arguments

- **x**: An object of class `crs`.
- **H**: An integer indicating the number of response-style-based clusters to display the correction result. If `H=NULL` and `cls.rs.vec=NULL`, `H` is set as `H=n`. If `H=NULL` but `cls.rs.vec!=NULL`, `H` is set as `H=max(cls.rs.vec)`. The default is `H=NULL`.
- **cls.rs.vec**: An integer vector of length `n` indicating response-style-based clusters for `n` respondents. If `cls.rs.vec=NULL` and `H!=NULL`, clusters are determined by k-means clustering of Beta. The default is `cls.rs.vec=NULL`.
- **...**: Additional arguments passed to `plot`.

Details

Correction results for each respondent are displayed. If either response-style-based clusters or the number of response-style-based clusters are specified, the correction results of response-style-based clusters are displayed.

References


See Also

- `ccrs`

Examples

```r
###data setting
n <- 30 ; m <- 10 ; H.true <- 2 ; K.true <- 2 ; q <- 5
datagene <- generate.rsdata(n=n,m=m,K.true=K.true,H.true=H.true,q=q,clustered.rs = TRUE)
##obtain n x m data matrix
X <- datagene$X
crdata.list <- create.ccrsdata(X,q=q)
ccrs.list <- correct.rs(crrsdata.list)
##You can check correction result using this \code{crs.plot} function.
plot(ccrs.list)

##You can also check correction result obtained by a simultaneous analysis of correction and content-based clustering.
##CCRS
lam <- 0.8 ; K <- 2
cccrs.list <- cccrs(ccrsdata.list,K=K,lam=lam)
##check correction result using this \code{ccrs.plot} function.
plot(cccrs.list$crs.list)
```
transformRSdata

Transform data by the estimated response function

Description
Transforms data matrix by estimated response functions.

Usage
transformRSdata(X, Beta=Beta, Mmat.q=Mmat.q)

Arguments
X
An n by m categorical data matrix.
Beta
An n by q-1 matrix of coefficients for response functions.
Mmat.q
A q by 3+1 matrix of I-spline basis functions, evaluated at the midpoints between boundaries.

Value
Returns a list with the following elements.
Y.hat
An n by m matrix of corrected data matrix.
MB
An n by q matrix of values of response functions evaluated at the midpoint between boundaries.
Index

* package
  ccrs-package, 2
* utility
  plot.crs, 7

ccrs, 2, 8
ccrs-package, 2
convert.X2F, 4
correct.rs, 3, 4, 6
create.ccrsdata, 4, 5, 5, 7

generate.rsdata, 6

plot, 8
plot.crs, 7

transformRSdata, 9