Package 'ExtMallows'

July 5, 2018

Type Package

Title An Extended Mallows Model and Its Hierarchical Version for Ranked Data Aggregation

Version 0.1.0

Date 2018-06-28

Description For multiple full/partial ranking lists, R package 'ExtMallows' can (1) detect whether the input ranking lists are over-correlated, and (2) use the Mallows model or extended Mallows model to integrate the ranking lists, and (3) use hierarchical extended Mallows model for rank integration if there are groups of over-correlated ranking lists.

Author Han Li, Minxuan Xu, Jun S. Liu and Xiaodan Fan

Maintainer Han Li <hli@szu.edu.cn>

Depends R (>= 3.1.0)

License GPL (>= 2)

NeedsCompilation no

Repository CRAN

Date/Publication 2018-07-05 15:30:10 UTC

R topics documented:

corrRankings *p value for measuring the correlation of pairwise rankings*

Description

It caclulates the p values that measure the correlation of pariwise rankings.

Usage

```
corrRankings(rankings)
```

Arguments

rankings	A n by m data frame, with each column representing a ranking list, which ranks the items from the most preferred to the least preferred. For missing items, use 0 to denote them.

Value

pair.pvalue	a symmetric matrix of p values, with the (i,j)-th element denoting the p value of the i,j-th rankings.

Note

Note that the input rankings should have at least 8 rankings. When constructing the samples of rescaled V distance for a given rank position, the number of samples should at least be 28 and the number of rankings that have items up to this position should account for at least 2/3 of the total number of rankings, otherwise the p value calculation stops at this position.

Author(s)

Han Li, Minxuan Xu, Jun S. Liu and Xiaodan Fan

References

An extended Mallows model for ranked data aggregation

Examples

```
data(simu3)
pvalue=corrRankings(rankings = simu3)

#threshold the p values

threshold=0.05
pvalue.trunc=ifelse(pvalue<=0.05, pvalue, 1)

#plot the p values
```

```
x=y=1:ncol(pvalue)
par(mfrow=c(1,2))
image(x, y, pvalue, xlab = NA, ylab = NA, sub = "rank coefficient")
image(x, y, pvalue.trunc, xlab = NA, ylab = NA, sub = "rank coefficient < 0.05")
```

EMM *An extended Mallows model for aggregating multiple ranking lists*

Description

It uses the extended Mallows model to aggregate multiple full/partial ranking lists.

Usage

```
EMM(rankings, initial.method, it.max)
```

Arguments

rankings	A n by m matrix, with each column representing a ranking list, which ranks the items from the most preferred to the least preferred. For missing items, use 0 to denote them.
initial.method	the method for initializing the value of pi0, with four options: mean, median, geometric and random (the mean of three randomly sampled ranking lists). By default, initial.method="mean".
it.max	the maximum number of iterations. By default, it.max=20.

Value

op.phi	optimal value of phi
op.omega	optimal value of omega
op.alpha	optimal value of alpha
op.pi0	optimal value of pi0, ranking the items from the most preferred to the least preferred
max.logL	maximum value of log-likelihood

Author(s)

Han Li, Minxuan Xu, Jun S. Liu and Xiaodan Fan

References

An extended Mallows model for ranked data aggregation

Examples

```
data(simu1)
res=EMM(rankings = simu1, initial.method = "mean", it.max = 20)
res$op.phi
res$op.omega
res$op.pi0
```

HEMM *A hierarchical extended Mallows model for aggregating multiple ranking lists*

Description

It uses the hierarchical extended Mallows model to aggregate multiple full/partial ranking lists.

Usage

```
HEMM(rankings, num.kappa, is.kappa.ranker, initial.method, it.max)
```

Arguments

rankings	A n by m matrix, with each column representing a ranking list, which ranks the items from the most preferred to the least preferred. For missing items, use 0 to denote them.
num.kappa	the number of over-correlated ranking groups
is.kappa.ranker	a list of over-correlated ranking groups, with the k-th element denoting the column numbers of the rankings that belong to the k-th group
initial.method	the method for initializing the value of pi0, with four options: mean, median, geometric and random (the mean of three randomly sampled ranking lists). By default, initial.method="mean".
it.max	the maximum number of iterations. By default, it.max=20.

Value

op.phi	optimal value of phi
op.phi1	optimal value of phi1, the phi value in over-correlated ranking groups
op.omega	optimal value of omega
op.alpha	optimal value of alpha
op.pi0	optimal value of pi0, ranking the items from the most preferred to the least preferred
op.kappa	optimal value of kappa, denoting the items from the most preferred to the least preferred
max.logL	maximum value of log-likelihood

Author(s)

Han Li, Minxuan Xu, Jun S. Liu and Xiaodan Fan

References

An extended Mallows model for ranked data aggregation

Examples

```
data(simu3)
res=HEMM(rankings = simu3, num.kappa = 2, is.kappa.ranker = list(1:5, 6:10),
    initial.method = "mean", it.max = 20)
res$op.phi
res$op.phi1
res$op.omega
res$op.pi0

data(NBArankings)
res=HEMM(rankings = NBArankings, num.kappa = 1, is.kappa.ranker = list(1:6),
    initial.method = "mean", it.max = 20)
res$op.omega
res$op.pi0
res$op.kappa
```

MM *The Mallows model for aggregating multiple ranking lists*

Description

It uses the Mallows model to aggregate multiple full/partial ranking lists.

Usage

```
MM(rankings, initial.method, it.max)
```

Arguments

`rankings`	A n by m matrix, with each column representing a ranking list, which ranks the items from the most preferred to the least preferred. For missing items, use 0 to denote them.
`initial.method`	the method for initializing the value of pi0, with four options: mean, median, geometric and random (the mean of three randomly sampled ranking lists). By default, initial.method="mean".
`it.max`	the maximum number of iterations. By default, it.max=20.

Value

op.phi	optimal value of phi
op.pi0	optimal value of pi0, ranking the items from the most preferred to the least preferred
max.logL	maximum value of log-likelihood

Author(s)

Han Li, Minxuan Xu, Jun S. Liu and Xiaodan Fan

References

Mallows, C. L. (1957). Non-null ranking models, Biometrika 44(1/2): 114-130.

Examples

```
data(simu1)
res=MM(rankings = simu1, initial.method = "mean", it.max = 20)
res$op.phi
res$op.pi0
```

NBArankings *A real example of rankings of NBA teams*

Description

This example is about aggregating the multiple rankings of NBA teams and was studied by Deng et al. (2014). They collected 34 rankings, including 6 professional rankings and 28 amateur rankings, for the 30 NBA teams in the 2011-2012 season. For the missing items in the partial rankings, we use number 0 to denote them.

Usage

```
data("NBArankings")
```

Format

A data frame with 30 observations on the following 34 variables.

V1 a factor with levels 76ers Bobcats Bucks Bulls Cavaliers Celtics Clippers Grizzlies Hawks Heat Hornets Jazz Kings Knicks Lakers Magic Mavericks Nets Nuggets Pacers Pistons Raptors Rockets Spurs Suns Thunder Timberwolves TrailBlazers Warriors Wizards

V2 a factor with levels 76ers Bobcats Bucks Bulls Cavaliers Celtics Clippers Grizzlies Hawks Heat Hornets Jazz Kings Knicks Lakers Magic Mavericks Nets Nuggets Pacers Pistons Raptors Rockets Spurs Suns Thunder Timberwolves TrailBlazers Warriors Wizards

V3 a factor with levels 76ers Bobcats Bucks Bulls Cavaliers Celtics Clippers Grizzlies Hawks Heat Hornets Jazz Kings Knicks Lakers Magic Mavericks Nets Nuggets Pacers Pistons Raptors Rockets Spurs Suns Thunder Timberwolves TrailBlazers Warriors Wizards

V4 a factor with levels 76ers Bobcats Bucks Bulls Cavaliers Celtics Clippers Grizzlies Hawks Heat Hornets Jazz Kings Knicks Lakers Magic Mavericks Nets Nuggets Pacers Pistons Raptors Rockets Spurs Suns Thunder Timberwolves TrailBlazers Warriors Wizards

V5 a factor with levels 76ers Bobcats Bucks Bulls Cavaliers Celtics Clippers Grizzlies Hawks Heat Hornets Jazz Kings Knicks Lakers Magic Mavericks Nets Nuggets Pacers Pistons Raptors Rockets Spurs Suns Thunder Timberwolves TrailBlazers Warriors Wizards

V6 a factor with levels 76ers Bobcats Bucks Bulls Cavaliers Celtics Clippers Grizzlies Hawks Heat Hornets Jazz Kings Knicks Lakers Magic Mavericks Nets Nuggets Pacers Pistons Raptors Rockets Spurs Suns Thunder Timberwolves TrailBlazers Warriors Wizards

V7 a factor with levels 0 Bulls Celtics Hawks Heat Lakers Pacers Spurs Thunder

V8 a factor with levels 0 Bulls Celtics Clippers Heat Knicks Lakers Spurs Thunder

V9 a factor with levels 0 Bulls Celtics Heat Knicks Lakers Mavericks Spurs Thunder

V10 a factor with levels 0 Bulls Celtics Clippers Heat Lakers Mavericks Spurs Thunder

V11 a factor with levels 0 Bulls Celtics Heat Knicks Lakers Nuggets Warriors Wizards

V12 a factor with levels 0 Bulls Celtics Clippers Heat Lakers Mavericks Spurs Thunder

V13 a factor with levels 0 Bulls Celtics Hornets Jazz Kings Lakers Magic Rockets

V14 a factor with levels 0 76ers Celtics Heat Kings Lakers Rockets Spurs Suns

V15 a factor with levels 0 Bulls Celtics Heat Lakers Mavericks Rockets Spurs Thunder

V16 a factor with levels 0 Celtics Hawks Heat Lakers Mavericks Raptors Spurs Thunder

V17 a factor with levels 0 76ers Celtics Heat Knicks Lakers Mavericks Nets Thunder

V18 a factor with levels 0 76ers Bulls Cavaliers Celtics Heat Lakers Mavericks Thunder

V19 a factor with levels 0 Bulls Heat Kings Lakers Rockets Spurs Suns Warriors

V20 a factor with levels 0 Bucks Celtics Heat Lakers Magic Mavericks Rockets Suns

V21 a factor with levels 0 Celtics Heat Kings Lakers Mavericks Spurs Suns Timberwolves

V22 a factor with levels 0 Celtics Heat Kings Lakers Spurs Suns Thunder Timberwolves

V23 a factor with levels 0 Bobcats Celtics Heat Lakers Mavericks Nuggets Spurs Suns

V24 a factor with levels 0 76ers Heat Knicks Lakers Pistons Rockets Spurs Wizards

V25 a factor with levels 0 76ers Celtics Hawks Heat Knicks Lakers Magic Thunder

V26 a factor with levels 0 Bulls Cavaliers Celtics Hawks Heat Knicks Lakers Rockets

V27 a factor with levels 0 76ers Clippers Lakers Magic Mavericks Pacers Raptors Warriors

V28 a factor with levels 0 76ers Bulls Celtics Heat Lakers Pistons Rockets Wizards

V29 a factor with levels 0 76ers Bulls Grizzlies Hawks Kings Knicks Nets Timberwolves

V30 a factor with levels 0 76ers Bucks Bulls Knicks Raptors Rockets Thunder Timberwolves

V31 a factor with levels 0 76ers Heat Lakers Magic Mavericks Pacers Pistons Suns

V32 a factor with levels 0 76ers Bulls Celtics Heat Knicks Lakers Magic Pacers

V33 a factor with levels 0 Clippers Heat Knicks Lakers Mavericks Nets Nuggets Wizards

V34 a factor with levels 0 Bulls Hawks Heat Jazz Knicks Nets Rockets Timberwolves

References

Deng, K., Han, S., Li, K. J. and Liu, J. S. (2014). Bayesian aggregation of order-based rank data, Journal of the American Statistical Association 109(507): 1023-1039.

Examples

```
data(NBArankings)
dim(NBArankings)
```

---

simu1 *Simulation data 1*

Description

This data set is simulated as described in the Simulation Study 1 of the reference. It is a 30 by 6 data frame, representing 6 independent top-30 partial rankings.

Usage

```
data("simu1")
```

Format

A data frame with 30 observations on the following 6 variables.

V1 a numeric vector

V2 a numeric vector

V3 a numeric vector

V4 a numeric vector

V5 a numeric vector

V6 a numeric vector

References

An extended Mallows model for ranked data aggregation

Examples

```
data(simu1)
dim(simu1)
```

simu2 *Simulation data 2*

Description

This data set is simulated as described in the Simulation Study 2 of the reference. It is a 40 by 6 data frame, representing 6 independent top-40 partial rankings.

Usage

```
data("simu2")
```

Format

A data frame with 40 observations on the following 6 variables.

`V1` a numeric vector

`V2` a numeric vector

`V3` a numeric vector

`V4` a numeric vector

`V5` a numeric vector

`V6` a numeric vector

References

An extended Mallows model for ranked data aggregation

Examples

```
data(simu2)
dim(simu2)
```

simu3 *Simulation data 3*

Description

This data set is simulated as described in the Simulation Study 3 of the reference. It is a 100 by 20 data frame, representing 20 full rankings. The columns 1-5 and the columns 6-10 represent two highly correlated ranking groups, respectively.

Usage

```
data("simu3")
```

Format

A data frame with 100 observations on the following 20 variables.

`V1` a numeric vector

`V2` a numeric vector

`V3` a numeric vector

`V4` a numeric vector

`V5` a numeric vector

`V6` a numeric vector

`V7` a numeric vector

`V8` a numeric vector

`V9` a numeric vector

`V10` a numeric vector

`V11` a numeric vector

`V12` a numeric vector

`V13` a numeric vector

`V14` a numeric vector

`V15` a numeric vector

`V16` a numeric vector

`V17` a numeric vector

`V18` a numeric vector

`V19` a numeric vector

`V20` a numeric vector

References

An extended Mallows model for ranked data aggregation

Examples

```
data(simu3)
dim(simu3)
```

Index