Package 'gCat'

June 6, 2022

Version 0.2

Date 2022-06-05

Title Graph-Based Two-Sample Tests for Categorical Data

Author Hao Chen and Nancy R. Zhang

Maintainer Hao Chen <hxchen@ucdavis.edu>

Depends R (>= 3.0.1)

Description These are two-sample tests for categorical data utilizing similarity information among the categories. They are useful when there is underlying structure on the categories.

License GPL (>= 2)

NeedsCompilation yes

Repository CRAN

Date/Publication 2022-06-06 18:30:02 UTC

R topics documented:

---

gCat   *Graph-based two-sample tests for categorical data*

Description

These are two-sample tests for categorical data utilizing similarity information among the categories. They are useful when there is underlying structure on the categories.

Author(s)

Hao Chen and Nancy R. Zhang

Maintainer: Hao Chen (hxchen@ucdavis.edu)

References

Chen, H. and Zhang, N.R. (2013) Graph-based tests for two-sample comparisons of categorical data. Statistica Sinica, 23, 1479-1503.

See Also

gcat.test

Examples

```
data(Example)
gcat.test(mycounts,mydist)
```

---

gcat.test   *Graph-based two-sample tests for categorical data*

Description

This function performs the two-sample tests for categorical data utilizing similarity information among the categories. You can either provide a distance matrix on the categories (through the "distmatrix" argument) or a similarity graph on the categories directly (through the "C0" argument) or both. The outputs of this function are the test statistic(s) and p-value(s).

Usage

```
gcat.test(counts, distmatrix=NULL, C0=NULL, method="C-uMST", Nperm=0)
```

Arguments

counts	It is a K by 2 matrix, where K is the number of categories. It specifies the counts in the K categories for the two samples.
distmatrix	A K by K matrix, which is the distance matrix on the categories. This needs to be specified if you include any of the four methods – "aMST", "uMST", "C-uMST" and "C-uNNB" – in the "method" argument.
C0	A similarity graph on the categories. It is a E by 2 matrix, where E is the number of edges in the graph. Each row in C0 corresponds to an edge in the graph, and the two numbers are the category indices connected by the edge. This needs to be specified if you include "RC0" or "TC0" in the "method" argument.

method	This argument specifies the test statistic(s) to be computed. It can be any combination of {"aMST", "C-uMST", "uMST", "C-uNNB", "RC0", "TC0"}. If you choose more than one method, use c(,) to combine them. For example: c("C-uMST", "uMST", "RC0"). The details of the statistics can be found in the paper: Chen, H. and Zhang, N.R. (2013) Graph-based tests for two-sample comparisons of categorical data. Statistica Sinica, 23, 1479-1503.
Nperm	Number of permutations in calculating the permutation p-value. This needs to be specified if the method is "aMST". For other methods, specifying this argument would provide in the result the permutation p-value in addition to the approximate p-value, which is calculated through asymptotic theories.

Examples

```
data(Example)
gcat.test(mycounts,mydist,myedge,method=c("aMST","C-uMST","uMST","C-uNNB","RC0","TC0"),Nperm=1000)
gcat.test(mycounts,mydist,method=c("C-uMST","uMST"))
gcat.test(mycounts,mydist)
gcat.test(mycounts,myedge,method="RC0")
```

---

mycounts	*Example raw counts*

Description

This is a toy example, which has 8 categories. This 8 by 2 matrix stores the number of occurrences in each category for the two samples.

---

mydist	*Example distance matrix*

Description

This is the distance matrix associated with the toy example having 8 categories.

---

myedge	*Example similarity graph*

Description

This is an example of a similarity graph associated with the toy example having 8 categories. This similarity graph has 12 edges.

Index