Package 'nFCA'

February 20, 2015

Type Package

Title Numerical Formal Concept Analysis for Systematic Clustering

Version 0.3

Date 2015-02-10

Author Junheng Ma, Jiayang Sun, and Guo-Qiang Zhang

Maintainer Junheng Ma <jxm216@case.edu>

Description Numerical Formal Concept Analysis (nFCA) is a modern unsupervised learning tool for analyzing general numerical data. Given input data, this R package nFCA outputs two nFCA graphs: a H-graph and an I-graph that reveal systematic, hierarchical clustering and inherent structure of the data.

License GPL-2

SystemRequirements Ruby, Graphviz

NeedsCompilation no

Repository CRAN

Date/Publication 2015-02-15 07:35:41

R topics documented:

`nfca` *Numerical Formal Concept Analysis for Systematic Clustering*

Description

The R function nfca() is an implementation of the numerical Formal Concept Analysis (nFCA), a modern unsupervised learning tool for analyzing general numerical data developed in [3]. nfca() provides two nFCA graphs: a H-graph and an I-graph that reveal systematic, hierarchical clustering and inherent structure of the data.

Usage

```
nfca(data, type = 0, method = 'hist', choice = 1, n = 30, alpha = 0.05)
```

Arguments

data	The input numerical data. It can be a correlation matrix, a distance matrix, or a general data matrix of n x p dimensions, where n is the number of subjects and p is the dimension of variables. If the input is a general data matrix, nfca is performed based on the correlation matrix computed from the data.
type	The type of input data. The default type=0 represents a correlation matrix, while type=1 represents a distance matrix and type=2 represents a general data matrix.
method	The method of choosing a sequence of thresholds to scale FCA in building nFCA. The default method='hist' implements the 'histogram' method, applicable to all data types, type=0, 1, or 2. The method='CI' uses the 'confidence interval' method, currently applicable when data type is a correlation matrix, i.e. type=0.
choice	The choice of how to choose the thresholds for the 'histogram' method. The default choice=1 chooses the thresholds automatically while choice=0 allows the user to choose thresholds based on histograms shown on the screen manually.
n	The sample size used to compute the correlation matrix if the input data is 0, ie. correlation matrix. Required only if the threshold selection method is 'confidence interval'. The default value is 30.
alpha	The significance level used for 'confidence interval' method. The default value is 0.05.

Details

Numerical Formal Concept Analysis (nFCA) combines the merit of statistics, formal concept analysis (FCA), and a graphical visualization tool (Graphviz) to analyze the clustering and inherent structure of data. Its output is a pair of nFCA graphs, H- and I-graphs. H-graph maps systematic relations of hierarchical clusters. I-graph is a directed acyclic graph (DAG) that complements the H-graph by revealing inherent structures and connections from one member to the relevant member of another cluster.

The nFCA package includes our main R code and a supporting program in Ruby that implements the faster concept analysis algorithm developed by Dr. Zhang's team (Troy et al. 2007). If needed, Ruby compiler can be downloaded from https://www.ruby-lang.org.

The two nFCA outcome files, Hgraph.dot and Igraph.dot, can be visualized using Graphviz. Graphviz is a standard, powerful graphic visualization software, available at http://www.graphviz.org/. We have tested selected versions of Graphviz. Versions 2.26, 2.30, 2.38 for Mac OS Lion and Window work with this package. Do not use version 2.28, which has a known bug. For further instructions on how to use Graphviz for nFCA, see the 'Value' below or: http://sr2c.case.edu/nfca, for detailed installation instructions and examples of figures from Graphviz.

Value

Hgraph.dot	a dot file containing systematic clustering result.

`Igraph.dot` a dot file containing inherited clustering information.

To visualize H-graph.dot and I-graph.dot in Graphviz, choose 'fdp' as the LAYOUT engine for the H-graph, and choose 'neato' for the I-graph. These selections can be done in GUI versions of Graphviz in Window or Mac. In a Mac running macport or Linux machine on which Graphviz is installed, use the following commands to generate graphics outside R:

fdp -Tpng Hgraph.dot -o Hgraph.png
neato -Tpng Igraph.dot -o Igraph.png

Author(s)

Junheng Ma, Jiayang Sun and Guo-Qiang Zhang

References

Troy, A. D., Zhang, G.-Q. and Tian, Y. (2007) Faster Concept Analysis. Proceedings of the 15th International Conference on Conceptual Structures (ICCS 2007), 4604, 206–219.

Ma, J. (2010) Contributions to Numerical Formal Concept Analysis, Bayesian Predictive Inference, and Sample Size Determination. PhD thesis, Case Western Reserve University. http://rave.ohiolink.edu/etdc/view?acc_num=case1285341426

Ma, J., Sun, J. and Zhang, G.-Q. (2014) Numerical Formal Concept Analysis (nFCA): a New Systematic Clustering Technique. Under review.

Examples

```
# View a build-in correlation matrix: nfca_example
data("nfca_example", package = "nFCA")
nfca_example

# 1. using the default 'histogram' method and choosing threshold
# automatically
nfca(data = nfca_example)

# 2. using 'confidence interval' method with sample size 30 and
# choosing threshold automatically
nfca(data = nfca_example, method = "CI")

# The output files Hgraph.dot and Igraph.dot from #1 and #2 can
# be visualized as H- and I-graphs in Graphviz. In this example,
# the I-graphs from both 'histogram' and 'confidence interval'
# methods are identical, while two H-graphs are consistent to
# each other.
```

nfca_example *Family Correlation Coefficient Data*

Description

This data is correlation coefficient matrix among family members. The family members are Husband(H), Wife(W), Son(S), Daughter(D), Husband's Father(HF), Husband's Mother(HM), Wife's Father(WF), Wife's Mother(WM), Husband's Friend(H1), Wife's Friend(W1), and Others(O).

Usage

```
nfca_example
```

Format

data frame

Source

Ma, J. (2010), Contributions to Numerical Formal Concept Analysis, Bayesian Predictive Inference, and Sample Size Determination, PhD thesis, Case Western Reserve University.
http://rave.ohiolink.edu/etdc/view?acc_num=case1285341426

References

Ma, J., Sun, J., and Zhang, G.-Q. (2014), Numerical Formal Concept Analysis (nFCA): a New Systematic Clustering Technique. Under review.

Index