

# Package ‘ConsensusClusterPlus’

March 26, 2013

**Type** Package

**Imports** Biobase, ALL, graphics, stats, utils

**Title** ConsensusClusterPlus

**Version** 1.10.0

**Date** 2011-5-11

**Author** Matt Wilkerson

**Maintainer** Matt Wilkerson <mwilkers@med.unc.edu>

**Description** algorithm for determining cluster count and membership by stability evidence in unsupervised analysis

**License** GPL version 2

**biocViews** Software, Bioinformatics, Clustering

## R topics documented:

|                                |          |
|--------------------------------|----------|
| ConsensusClusterPlus . . . . . | 1        |
| <b>Index</b>                   | <b>4</b> |

---

|                      |                                 |
|----------------------|---------------------------------|
| ConsensusClusterPlus | <i>run ConsensusClusterPlus</i> |
|----------------------|---------------------------------|

---

## Description

ConsensusClusterPlus function for determining cluster number and class membership by stability evidence. calcICL function for calculating cluster-consensus and item-consensus.

## Usage

```
ConsensusClusterPlus(  
d=NULL, maxK = 3, reps=10, pItem=0.8, pFeature=1, clusterAlg="hc",title="untitled_consensus_cluster",  
innerLinkage="average", finalLinkage="average", distance="pearson", ml=NULL,  
tmyPal=NULL,seed=NULL,plot=NULL,writeTable=FALSE,weightsItem=NULL,weightsFeature=NULL,ver  
calcICL(res,title="untitled_consensus_cluster",plot=NULL,writeTable=FALSE)
```

**Arguments**

|                |   |
|----------------|---|
| d              | matrix where columns=items/samples and rows are features. For example, a gene expression matrix of genes in rows and microarrays in columns. OR ExpressionSet object. |
| maxK           | integer value. maximum cluster number to evaluate.  |
| reps           | integer value. number of subsamples.  |
| pItem          | numerical value. proportion of items to sample.   |
| pFeature       | numerical value. proportion of features to sample.  |
| clusterAlg     | character value. cluster algorithm. "hc" heirarchical (hclust) or "km" for k-means. See Note.   |
| title          | character value for output directory. Directory is created only if plot is not NULL or writeTable is TRUE. This title can be an absolute or relative path.            |
| innerLinkage   | heirarchical linkage method for subsampling.  |
| finalLinkage   | heirarchical linkage method for consensus matrix.   |
| distance       | character value. sample distance measures: "pearson", "spearman", or "euclidean".   |
| ml             | optional. prior result, if supplied then only do graphics and tables.   |
| tmyPal         | optional character vector of colors for consensus matrix  |
| seed           | optional numerical value. sets random seed for reproducible results.  |
| plot           | character value. NULL - print to screen, 'pdf', 'png'.  |
| writeTable     | logical value. TRUE - write output and log to csv.  |
| weightsItem    | optional numerical vector. weights to be used for sampling items.   |
| weightsFeature | optional numerical vector. weights to be used for sampling features.  |
| res            | result of consensusClusterPlus.   |
| verbose        | boolean. If TRUE, print messages to the screen to indicate progress. This is useful for large datasets.   |

**Details**

ConsensusClusterPlus implements the Consensus Clustering algorithm of Monti, et al (2003) and extends this method with new functionality and visualizations. Its utility is to provide quantitative stability evidence for determining a cluster count and cluster membership in an unsupervised analysis.

ConsensusClusterPlus takes a numerical data matrix of items as columns and rows as features. This function subsamples this matrix according to pItem, pFeature, weightsItem, and weightsFeature, and clusters the data into 2 to maxK clusters by clusterArg clusteringAlgorithm. Agglomerative heirarchical (hclust) and kmeans clustering are supported by an option see above. For users wishing to use a different clustering algorithm for which many are available in R, one can supply their own clustering algorithm as a simple programming hook - see the second commented-out example that uses divisive heirarchical clustering.

For a detailed description of usage, output and images, see the vignette by: openVignette().

**Value**

ConsensusClusterPlus returns a list of length maxK. Each element is a list containing consensusMatrix (numerical matrix), consensusTree (hclust), consensusClass (consensus class assignments). ConsensusClusterPlus also produces images.

calcICL returns a list of two elements clusterConsensus and itemConsensus corresponding to cluster-consensus and item-consensus. See Monti, et al (2003) for formulas.

**Author(s)**

Matt Wilkerson mwilkers@med.unc.edu

**References**

Monti, S., Tamayo, P., Mesirov, J., Golub, T. (2003) Consensus Clustering: A Resampling-Based Method for Class Discovery and Visualization of Gene Expression Microarray Data. *Machine Learning*, 52, 91-118.

**Examples**

```
## obtain gene expression data
library(Biobase)
data(geneData)
d=geneData
#median center genes
d = sweep(d,1, apply(d,1,median))

## run consensus cluster
rcc = ConsensusClusterPlus(d,maxK=4,rep=100,pItem=0.8,pFeature=1,title="example")

## ICL
resICL = calcICL(rcc,title="example")

##example of programming hook for clusterAlg:
#library(cluster)
#dianaHook = function(this_dist,k){
#  tmp = diana(this_dist,diss=TRUE)
#  assignment = cutree(tmp,k)
#  return(assignment)
#}
#ConsensusClusterPlus(d,maxK=6,rep=25,pItem=0.8,pFeature=1,title="example",plot="png",clusterAlg="dianaHook")
```

# Index

## \*Topic **methods**

ConsensusClusterPlus, [1](#)

calcICL (ConsensusClusterPlus), [1](#)

ConsensusClusterPlus, [1](#)