

Package ‘scSorter’

September 1, 2020

Type Package

Title Implementation of 'scSorter' Algorithm

Version 0.0.1

Description Implements the algorithm described in

Guo, H., and Li, J. (Not yet published), “scSorter: assigning cells to known cell types according to known marker genes”.

Clusters cells to known cell types based on marker genes specified for each cell type.

Depends R (>= 3.6.0)

Imports stats (>= 3.6.0)

License GPL-3

LazyData true

RoxygenNote 7.1.0

Suggests knitr, rmarkdown, Seurat

VignetteBuilder knitr

NeedsCompilation no

Author Hongyu Guo [aut],
Jun Li [aut, cre]

Maintainer Jun Li <jun.li@nd.edu>

Repository CRAN

Date/Publication 2020-09-01 14:20:02 UTC

R topics documented:

cost_func	2
data_preprocess	2
design_matrix_builder	3
scSorter	3
update_C	4
update_func	5
update_mu	6

Index	7
--------------	----------

`cost_func`*Cost Function*

Description

Calculates the cost.

Usage

```
cost_func(dat, clus, mu, designmat)
```

Arguments

<code>dat</code>	A matrix of input data.
<code>clus</code>	A vector of predicted cell types.
<code>mu</code>	Parameter estimates from <code>update_mu</code> .
<code>designmat</code>	An indicator variable matrix records specified marker genes of each cell type.

`data_preprocess`*Preprocess Data*

Description

This function validates and preprocesses the input data for the downstream analysis.

Usage

```
data_preprocess(expr, anno_processed)
```

Arguments

<code>expr</code>	A matrix of input data. Each row represents a gene and each column represents a cell.
<code>anno_processed</code>	A list of processed annotation information that consists of the design matrix and the weight matrix for marker genes.

Value

A list contains processed expression matrix, design matrix, and weight matrix.

design_matrix_builder	<i>Design Matrix Builder</i>
-----------------------	------------------------------

Description

Builds the design matrix required by update_func based on user input.

Usage

```
design_matrix_builder(anno, weight)
```

Arguments

anno	A matrix or data frame that contains marker genes specified for cell types of interest.
weight	The default weight assigned to marker genes.

Value

A list contains processed design matrix and weight matrix.

scSorter	<i>scSorter</i>
----------	-----------------

Description

This is the main function that implements the scSorter method.

Usage

```
scSorter(  
  expr,  
  anno,  
  default_weight = 2,  
  n_start = 10,  
  alpha = 0,  
  u = 0.05,  
  max_iter = 100,  
  setseed = 0  
)
```

Arguments

expr	A matrix of the input expression data. Each row represents a gene and each column represents a cell. Each row of this matrix should be named by the gene name it represents.
anno	A matrix or data frame that contains marker genes specified for cell types of interest. It should contain three columns named "Type", "Marker", and "Weight" that records the name and weight of marker genes specified for each cell type. "Weight" column is optional. If it is not specified, the default_weight will be applied to all marker genes.
default_weight	The default weight assigned to marker genes. The default value is 2.
n_start	The number of possible cluster initializations. The default value is 10.
alpha	The parameter determines the cutoff whether the cell type of a cell should be considered as undecided during unknown cell calling. The default value is 0.
u	The parameter determines whether undecided cells are further processed. The default value is 0.05.
max_iter	The maximum number of iterations for the algorithm to update parameters. The default value is 100.
setseed	Random seed for cluster initialization. The default value is 0.

Value

A list contains the elements: Pred_Type: The predicted cell types. Pred_param: The parameter estimates of mu and delta.

Examples

```
load(system.file('extdata', 'example_data.RData', package = 'scSorter'))
result = scSorter(expr, anno)
misclassification_rate = 1 - mean(result$Pred_Type == true_type)
table(result$Pred_Type, true_type)
```

update_C	<i>Update Cluster</i>
----------	-----------------------

Description

Updates cluster assignments based on center estimates from update_mu

Usage

```
update_C(dat, mu_mat, designmat)
```

Arguments

dat	A matrix of input data.
mu_mat	Center estimates from update_mu
designmat	An indicator variable matrix records specified marker genes of each cell type.

update_func	<i>Update Function</i>
-------------	------------------------

Description

Implements the scSorter method by iteratively running update_mu and update_C.

Usage

```
update_func(
  dat,
  design_mat,
  weightmat,
  unknown_threshold1 = 0,
  unknown_threshold2 = 0.05,
  max_iter = 100
)
```

Arguments

dat	A matrix of input data.
design_mat	An indicator variable matrix records specified marker genes of each cell type.
weightmat	A matrix of weights assigned to each marker gene.
unknown_threshold1	The parameter determines undecided cells cutoff. The default value is 0.
unknown_threshold2	The parameter determines whether undecided cells are further processed. The default value is 0.05.
max_iter	The maximum number of iterations for the algorithm to update parameters. The default value is 100.

Value

A list contains parameter estimates, type assignments, and the corresponding cost.

`update_mu`*Mu Update*

Description

Solves mu and delta given sample cluster assignment.

Usage

```
update_mu(dat, designmat, clus)
```

Arguments

<code>dat</code>	A matrix of input data.
<code>designmat</code>	An indicator variable matrix records marker genes of each pre-specified cell type.
<code>clus</code>	A vector of cluster assignment.

Value

A matrix of parameter estimates.

Index

cost_func, [2](#)

data_preprocess, [2](#)

design_matrix_builder, [3](#)

scSorter, [3](#)

update_C, [4](#)

update_func, [5](#)

update_mu, [6](#)