Package 'scMOO'

December 9, 2021

Type Package

Title Accurate gene expression recovery for single-cell RNA sequencing via multi-objective optimization

Version 1.1

Author Ke Jin, Xiao-Fei Zhang

Maintainer Ke Jin <kej13@mails.ccnu.edu.cn>

Description scMOO assumes that the data is a combination of three types of latent structures, horizontal structure (genes are similar with each other), vertical structure (cells are similar with each other), and low-rank structure. The combination weights and latent structures are learned using multi-objective optimization, and the weighted average of the observed data and the imputation results learned from the three types of structures is considered as the final result.

License GPL-2

Encoding UTF-8

LazyData true

RoxygenNote 6.1.1

Depends R (>= 3.5.0)

Imports keras, tensorflow, rsvd

Suggests knitr, rmarkdown

VignetteBuilder knitr

R topics documented:

PBMC_CL *Peripheral blood mononuclear cells (PBMC)*

Description

This is the PBMC dataset sequenced using platform CEL-Seq2 (GSE132044), thus we rename it as "PBMC_CL".

Usage

```
PBMC_CL
```

Format

An object of class `matrix` with 2000 rows and 243 columns.

Details

We download the preprocessed data from https://doi.org/10.5281/zenodo.3357167.

The data contains 20,041 genes and 243 cells with true cell labels. We then select 2,000 highly variable genes using Seurat v3.2 before imputation.

Author(s)

Ke Jin, <kej13@mails.ccnu.edu.cn>

References

Ding, J. et al (2020). Systematic comparison of single-cell and single-nucleus rna-sequencing methods. *Nat. Biotechnol.*, 38:737-746.

Examples

```
data("PBMC_CL")
```

scMOO *use scMOO to impute dropout values in scRNA-seq data*

Description

use scMOO to impute dropout values in scRNA-seq data

Usage

```
scMOO(Y.count, W = NULL, K = 0, q = 10, percent = 0.05,
  lambda1 = NULL, lambda2 = NULL, lambda3 = 1e+10, alpha = NULL,
  MAX.ITER = 4, ABSTOL = 0.001, learning.rate = 1e-04,
  epochs = 100, verbose = TRUE, estimates.only = FALSE)
```

Arguments

W	W_gc is set to the ratio of the number of non-zero elements in Y to the number of zero elements in Y if Y_gc = 0, and to 1 otherwise.
q	The parameter for number of additional power iterations.
percent	The expression count matrix is preprocessed by filtering out the genes expressed in at most percent*100% of the cells.
lambda1	Tuning parameter for entropy regularizer.
lambda2	Tuning parameter to facilitate feature selection and regularization.
lambda3	Tuning parameter to penalize the diagonal elements of the parameter to eliminate the trivial solution of representing an expression level as a linear combination of itself.
alpha	Tuning parameter to balance the error between the imputed and observed data and the error of model fitting data.
ABSTOL	Absolute tolerance of circulation.
epochs	The number of the entire training set going through the entire network.
verbose	Whether to output the value of metrics at the end of each epoch. Default is TRUE.
Y_count	An expression count matrix. The rows correspond to genes and the columns correspond to cells. Can be sparse.
k	The rank of the low-rank approximation matrix.
MAX_ITER	Maximum iteration of scMOO.
learning_rate	A hyper-parameter that controls the speed of adjusting the weights of the network with respect to the loss gradient.

Value

If 'estimates.only = TRUE', then a matrix of scMOO estimates.

If 'estimates.only = FALSE', a list with the following components

estimate	Recovered (normalized) expression.
size.factor	Size factor used for normalization.
pred.time	Total time for scMOO estimation.
alpha	Tuning parameter to balance the error between the imputed and observed data and the error of model fitting data.
w	The combination weights W1, W2 and W3.

Author(s)

Ke Jin, <kej13@mails.ccnu.edu.cn>

Examples

```
data("PBMC_CL")

result <- scMOO(PBMC_CL, percent = 0, estimates.only = TRUE)
```

Index