## Package 'scPOEM'

September 25, 2025

Author Yuntong Hou [aut, cre] (<a href="https://orcid.org/0009-0005-0587-4692">https://orcid.org/0009-0005-0587-4692</a>), Yan Zhong [aut, ctb] (<a href="https://orcid.org/0000-0003-2412-043X">https://orcid.org/0000-0003-2412-043X</a>),

Type Package

Version 0.1.3

Yeran Chen [ctb],

Title Single-Cell Meta-Path Based Omic Embedding

```
Youshi Chang [ctb],
      Yongjian Yang [ctb] (<a href="https://orcid.org/0000-0002-4135-5014">https://orcid.org/0000-0002-4135-5014</a>),
      Xinyue Zheng [ctb],
      James Cai [ctb] (<https://orcid.org/0000-0002-8081-6725>)
Description Provide a workflow to jointly embed chromatin accessibility peaks and ex-
      pressed genes into a shared low-dimensional space using paired single-cell ATAC-seq (scATAC-
      seq) and single-cell RNA-seq (scRNA-seq) data. It integrates regulatory relation-
      ships among peak-peak interactions (via 'Cicero'), peak-gene interactions (via Lasso, random for-
      est, and XGBoost), and gene-gene interactions (via principal component regression). With the in-
      put of paired scATAC-seq and scRNA-seq data matrices, it assigns a low-dimensional fea-
      ture vector to each gene and peak. Additionally, it supports the reconstruction of gene-gene net-
      work with low-dimensional projections (via epsilon-NN) and then the comparison of the net-
      works of two conditions through manifold alignment implemented in 'scTenifold-
      Net'. See <doi:10.1093/bioinformatics/btaf483> for more details.
URL https://github.com/Houyt23/scPOEM
BugReports https://github.com/Houyt23/scPOEM/issues
License GPL (>= 2)
Encoding UTF-8
Roxygen list(markdown = TRUE)
Imports
      methods, utils, stats, foreach (>= 1.5.2), doParallel (>= 1.0.17), tictoc (>= 1.2.1), Matrix (>= 1.6-
      8), xgboost (>= 1.7.10), reticulate, stringr, magrittr, scTenifoldNet, VGAM (>= 1.1-
      13), Biobase (>= 2.66.0), BiocGenerics (>= 0.52.0), monocle (>= 2.34.0), cicero (>= 1.24.0)
Depends R (>= 4.1.0)
RoxygenNote 7.3.2
```

2 align\_embedding

## **Contents**

align_embedding	2
eNN	4
example_data_compare	4
example_data_single	5
GGN	5
PGN_Lasso	6
PGN_RF	7
PGN_XGBoost	9
pg_embedding	10
PPN	11
scPOEM	13
	16
	10

align\_embedding

Gene Network Reconstruction and Alignment

## Description

Reconstruct gene networks via epsilon-NN and compare conditions using manifold alignment implemented in scTenifoldNet.

## Usage

Index

```
align_embedding(
  gene_data1,
  gene_node1,
  E1,
  gene_data2,
  gene_node2,
  E2,
  dirpath = tempdir(),
  save_file = TRUE,
  d = 100
)
```

## **Arguments**

gene_data1	The information for genes in state1, must have a col names "gene_name".
gene_node1	Gene ids that are associated with other peaks or genes in state1.
E1	Embedding representations of peaks and genes in state1.
gene_data2	The information for genes in state2, must have a col names "gene_name".
gene_node2	Gene ids that are associated with other peaks or genes in state2.
E2	Embedding representations of peaks and genes in state2.
dirpath	The folder path to read or write file
save_file	Logical, whether to save the output to a file.
d	The dimension of latent space.

align\_embedding 3

#### Value

A list containing the following elements:

E\_g2 Low-dimensional embedding representations of genes under the two conditions. common\_genes Genes shared between both conditions and used in the analysis. diffRegulation A list of differential regulatory information for each gene.

```
library(scPOEM)
library(monocle)
dirpath <- "./example_data"</pre>
# Download compare mode example data
data(example_data_compare)
data_S1 <- example_data_compare$S1</pre>
data_S2 <- example_data_compare$S2</pre>
gg_net1 <- GGN(data_S1$Y, file.path(dirpath, "compare/S1"), save_file=FALSE)</pre>
pp_net1 <- PPN(data_S1$X, data_S1$peak_data, data_S1$cell_data,</pre>
                data_S1$genome, file.path(dirpath, "compare/S1"), save_file=FALSE)
net_Lasso1 <- PGN_Lasso(data_S1$X, data_S1$Y,</pre>
                         data_S1$gene_data, data_S1$neibor_peak,
                         file.path(dirpath, "compare/S1"), save_file=FALSE)
net_RF1 <- PGN_RF(data_S1$X, data_S1$Y, data_S1$gene_data,</pre>
                  data_S1$neibor_peak, file.path(dirpath, "compare/S1"), save_file=FALSE)
net_XGB1 <- PGN_XGBoost(data_S1$X, data_S1$Y,</pre>
                         data_S1$gene_data, data_S1$neibor_peak,
                         file.path(dirpath, "compare/S1"), save_file=FALSE)
pg_net_list1 <- list(net_Lasso1, net_RF1, net_XGB1)</pre>
E_result_S1 <- pg_embedding(gg_net1, pp_net1, pg_net_list1,</pre>
                file.path(dirpath, "compare/S1"), save_file=FALSE)
gg_net2 <- GGN(data_S2$Y, file.path(dirpath, "compare/S2"), save_file=FALSE)</pre>
pp_net2 <- PPN(data_S2$X, data_S2$peak_data,</pre>
                data_S2$cell_data, data_S2$genome,
                file.path(dirpath, "compare/S2"), save_file=FALSE)
net_Lasso2 <- PGN_Lasso(data_S2$X, data_S2$Y,</pre>
                         data_S2$gene_data, data_S2$neibor_peak,
                         file.path(dirpath, "compare/S2"), save_file=FALSE)
net_RF2 <- PGN_RF(data_S2$X, data_S2$Y, data_S2$gene_data,</pre>
                  data_S2$neibor_peak, file.path(dirpath, "compare/S2"), save_file=FALSE)
net_XGB2 <- PGN_XGBoost(data_S2$X, data_S2$Y,</pre>
                         data_S2$gene_data, data_S2$neibor_peak,
                         file.path(dirpath, "compare/S2"), save_file=FALSE)
pg_net_list2 <- list(net_Lasso2, net_RF2, net_XGB2)</pre>
E_result_S2 <- pg_embedding(gg_net2, pp_net2, pg_net_list2,</pre>
                file.path(dirpath, "compare/S2"), save_file=FALSE)
compare_result <- align_embedding(data_S1$gene_data,</pre>
                                    E_result_S1$gene_node,
                                    E_result_S1$E,
                                    data_S2$gene_data,
                                    E_result_S2$gene_node,
                                    E_result_S2$E,
                                    file.path(dirpath, "compare/compare"),
                                    save_file=FALSE)
```

eNN

Network Reconstruction via epsilon-NN

### **Description**

Reconstruction of gene-gene network via low-dimentional projections (via epsilon-NN).

## Usage

eNN(E\_g)

### **Arguments**

E\_g

Embedding representations of genes.

#### Value

The epsilon-NN network.

example\_data\_compare

Example Input Data for Compare Mode Analysis

## **Description**

A list containing example single-cell multi-omics data used in "compare" mode of the scP0EM package.

## Usage

```
data(example_data_compare)
```

#### **Format**

A named list of length 2. Each element is itself a named list with the following components:

X The scATAC-seq data, sparse matrix.

Y The scRNA-seq data, sparse matrix.

peak\_data A data.frame containing peak information.

gene\_data A data.frame containing gene information (must contain column "gene\_name").

cell\_data A data.frame containing cell metadata.

neibor\_peak The peak IDs within a certain range of each gene, must have cols c("gene\_name", "start\_use", "end\_use"). The id numbers in "start\_use" and "end\_use" are start from 0. genome The genome length for the species.

```
data(example_data_compare)
```

example\_data\_single 5

```
example_data_single I
```

Example Input Data for Single Mode Analysis

## **Description**

A list containing example single-cell multi-omics data used in "single" mode of the scP0EM package.

## Usage

```
data(example_data_single)
```

#### **Format**

```
A named list with 7 elements:
```

```
X The scATAC-seq data, sparse matrix.
```

Y The scRNA-seq data, sparse matrix.

peak\_data A data.frame containing peak information.

gene\_data A data.frame containing gene information (must contain column "gene\_name").

cell\_data A data.frame containing cell metadata.

neibor\_peak The peak IDs within a certain range of each gene, must have cols  $c("gene_name", "start_use", "end_use")$ . The id numbers in "start\_use" and "end\_use" are start from 0.

genome The genome length for the species.

### **Examples**

```
data(example_data_single)
```

GGN

Construct Gene-Gene Network

## Description

Construct the gene-gene network via principle component regression.

#### Usage

```
GGN(
   Y,
   dirpath = tempdir(),
   count_device = 1,
   nComp = 5,
   rebuild_GGN = TRUE,
   save_file = TRUE,
   python_env = "scPOEM_env"
)
```

6 PGN\_Lasso

## **Arguments**

Y The scRNA-seq data, sparse matrix. dirpath The folder path to read or write file.

count\_device The number of cpus used to train the Lasso model.

nComp The number of PCs used for regression

rebuild\_GGN Logical. Whether to rebuild the gene-gene network (GGN) from scratch. If

FALSE, the function will attempt to read from GGN. mtx under dirpath/test in

single mode or dirpath/state\_name/test in compare mode.

save\_file Logical, whether to save the output to a file.

python\_env Name or path of the Python environment to be used.

#### Value

The GGN network.

#### **Examples**

PGN\_Lasso

Peak-Gene Network via Lasso

#### **Description**

Construct the peak-gene network via Lasso.

### Usage

```
PGN_Lasso(
   X,
   Y,
   gene_data,
   neibor_peak,
   dirpath = tempdir(),
   count_device = 1,
   rebuild_PGN_Lasso = TRUE,
   save_file = TRUE
)
```

7 PGN\_RF

#### **Arguments**

Χ The scATAC-seq data, sparse matrix. The scRNA-seq data, sparse matrix. gene\_data The information for genes, must have a col names "gene\_name". The peak IDs within a certain range of each gene, must have cols c("gene\_name", neibor\_peak "start\_use", "end\_use"). The id numbers in "start\_use" and "end\_use" are start dirpath The folder path to read or write file. count\_device The number of cpus used to train the Lasso model. rebuild\_PGN\_Lasso Logical. Whether to rebuild the peak-gene network via Lasso from scratch. If FALSE, the function will attempt to read from PGN\_Lasso.mtx under

dirpath/test in single mode or dirpath/state\_name/test in compare mode.

save\_file Logical, whether to save the output to a file.

### Value

The PGN\_Lasso network.

## **Examples**

```
library(scPOEM)
dirpath <- "./example_data"</pre>
# Download single mode example data
data(example_data_single)
# Construct PGN net via Lasso.
net_Lasso <- PGN_Lasso(example_data_single$X,</pre>
                        example_data_single$Y,
                        example_data_single$gene_data,
                        example_data_single$neibor_peak,
                        file.path(dirpath, "single"),
                        save_file=FALSE)
```

PGN\_RF

Peak-Gene Network via Random Forest

#### **Description**

Construct the peak-gene network via random forest.

#### Usage

```
PGN_RF(
 Χ,
  Υ,
 gene_data,
 neibor_peak,
 dirpath = tempdir(),
```

8 PGN\_RF

```
count_device = 1,
rebuild_PGN_RF = TRUE,
save_file = TRUE,
seed = NULL,
python_env = "scPOEM_env"
)
```

### **Arguments**

X The scATAC-seq data, sparse matrix.

Y The scRNA-seq data, sparse matrix.

gene\_data The information for genes, must have a col names "gene\_name".

neibor\_peak The peak IDs within a certain range of each gene, must have cols c("gene\_name",

"start\_use", "end\_use"). The id numbers in "start\_use" and "end\_use" are start

from 0.

dirpath The folder path to read or write file.

count\_device The number of cpus used to train the Lasso model.

rebuild\_PGN\_RF Logical. Whether to rebuild the peak-gene network via random forest from

scratch. If FALSE, the function will attempt to read from PGN\_RF.mtx under dirpath/test in single mode or dirpath/state\_name/test in compare

mode.

save\_file Logical, whether to save the output to a file.

seed An integer specifying the random seed to ensure reproducible results.

python\_env Name or path of the Python environment to be used.

## Value

The PGN\_RF network.

PGN\_XGBoost 9

PGN	XGBoost

Peak-Gene Network via XGBoost

## **Description**

Construct the peak-gene network via XGBoost.

### Usage

```
PGN_XGBoost(
   X,
   Y,
   gene_data,
   neibor_peak,
   dirpath = tempdir(),
   count_device = 1,
   rebuild_PGN_XGB = TRUE,
   save_file = TRUE
)
```

## Arguments

X The scATAC-seq data, sparse matrix.
Y The scRNA-seq data, sparse matrix.
gene\_data The information for genes, must have a col name

gene\_data The information for genes, must have a col names "gene\_name".

neibor\_peak The peak IDs within a certain range of each gene, must have cols column.

The peak IDs within a certain range of each gene, must have cols c("gene\_name",

"start\_use", "end\_use"). The id numbers in "start\_use" and "end\_use" are start

from 0.

dirpath The folder path to read or write file.

count\_device The number of cpus used to train the Lasso model.

rebuild\_PGN\_XGB

Logical. Whether to rebuild the peak-gene network via XGBoost from scratch.

If FALSE, the function will attempt to read from PGN\_XGB.mtx under

 $\verb|dirpath/test| in single mode| or \verb|dirpath/state_name/test| in compare mode.$ 

save\_file Logical, whether to save the output to a file.

#### Value

The PGN\_XGBoost network.

10 pg\_embedding

```
example_data_single$neibor_peak,
file.path(dirpath, "single"),
save_file=FALSE)
```

pg\_embedding

Co-embeddings of Peaks and Genes.

## **Description**

Learn the low-dimensional representations for peaks and genes with a meta-path based method.

#### Usage

```
pg_embedding(
  gg_net,
  pp_net,
  pg_net_list,
  dirpath = tempdir(),
  relearn_pg_embedding = TRUE,
  save_file = TRUE,
  d = 100,
  numwalks = 5,
  walklength = 3,
  epochs = 100,
  neg\_sample = 5,
  batch_size = 32,
  weighted = TRUE,
  exclude_pos = FALSE,
  seed = NULL,
  python_env = "scPOEM_env"
)
```

#### **Arguments**

gg\_net The gene-gene network. The peak-peak network. pp\_net A list of peak-gene networks, constructed via different methods. pg\_net\_list dirpath The folder path to read or write file. relearn\_pg\_embedding Logical. Whether to relearn the low-dimensional representations for peaks and genes from scratch. If FALSE, the function will attempt to read from node\_embeddings.mtx, node\_used\_peak.csv, node\_used\_gene.csv under dirpath/embedding in single mode or dirpath/state\_name/embedding in compare mode. save\_file Logical, whether to save the output to a file. Dimension of the latent space. Default is 100. numwalks Number of random walks per node. Default is 5.

PPN 11

walklength Length of walk depth. Default is 3.

epochs Number of training epochs. Default is 100.

neg\_sample Number of negative samples per positive sample. Default is 5.

batch\_size Batch size for training. Default is 32.

weighted Whether the sampling network is weighted. Default is TRUE.

exclude\_pos Whether to exclude positive samples from negative sampling. Default is FALSE.

seed An integer specifying the random seed to ensure reproducible results.

python\_env Name or path of the Python environment to be used.

#### Value

A list containing the following:

E Low-dimensional representations of peaks and genes peak\_node Peak ids that are associated with other peaks or genes. gene\_node Gene ids that are associated with other peaks or genes.

## **Examples**

```
library(scPOEM)
library(monocle)
dirpath <- "./example_data"</pre>
# Download single mode example data
data(example_data_single)
gg_net <- GGN(example_data_single$Y,</pre>
              file.path(dirpath, "single"),
              save_file=FALSE)
pp_net <- PPN(example_data_single$X, example_data_single$peak_data,</pre>
              example_data_single$cell_data, example_data_single$genome,
              file.path(dirpath, "single"), save_file=FALSE)
net_Lasso <- PGN_Lasso(example_data_single$X, example_data_single$Y,</pre>
                        example_data_single$gene_data, example_data_single$neibor_peak,
                        file.path(dirpath, "single"), save_file=FALSE)
net_RF <- PGN_RF(example_data_single$X, example_data_single$Y,</pre>
                  example_data_single$gene_data, example_data_single$neibor_peak,
                  file.path(dirpath, "single"), save_file=FALSE)
net_XGB <- PGN_XGBoost(example_data_single$X, example_data_single$Y,</pre>
                        example_data_single$gene_data, example_data_single$neibor_peak,
                        file.path(dirpath, "single"), save_file=FALSE)
E_result <- pg_embedding(gg_net, pp_net, list(net_Lasso, net_RF, net_XGB),</pre>
                          file.path(dirpath, "single"), save_file=FALSE)
```

PPN

Construct Peak-Peak Network

#### **Description**

Construct peak-peak network.

12 PPN

### Usage

```
PPN(
    X,
    peak_data,
    cell_data,
    genome,
    dirpath = tempdir(),
    rebuild_PPN = TRUE,
    save_file = TRUE,
    seed = NULL
)
```

## Arguments

X	The scATAC-seq data, sparse matrix.
peak_data	The information for peaks, must have a col names "peak_name".
cell_data	The information for cells, must have a col names "cell_name".
genome	The genome length for the species.
dirpath	The folder path to read or write file.
rebuild_PPN	Logical. Whether to rebuild the peak-peak network (PPN) from scratch. If FALSE, the function will attempt to read from PPN.mtx under dirpath/test in single mode or dirpath/state_name/test in compare mode.
save_file	Logical, whether to save the output to a file.
seed	An integer specifying the random seed to ensure reproducible results.

## Value

The PPN network.

scPOEM 13

scP0EM

Main Function.

## **Description**

This function takes paired single-cell ATAC-seq (scATAC-seq) and RNA-seq (scRNA-seq) data to embed peaks and genes into a shared low-dimensional space. It integrates regulatory relationships from peak-peak interactions (via Cicero), peak-gene interactions (via Lasso, random forest, and XGBoost), and gene-gene interactions (via principal component regression). Additionally, it supports gene-gene network reconstruction using epsilon-NN projections and compares networks across conditions through manifold alignment (scTenifoldNet).

## Usage

```
scPOEM(
 mode = c("single", "compare"),
  input_data,
 dirpath = tempdir(),
  count_device = 1,
 nComp = 5,
  seed = NULL,
 numwalks = 5.
 walklength = 3,
 epochs = 100,
 neg\_sample = 5,
 batch_size = 32,
 weighted = TRUE,
  exclude_pos = FALSE,
  d = 100,
  rebuild_GGN = TRUE,
  rebuild_PPN = TRUE,
  rebuild_PGN_Lasso = TRUE,
  rebuild_PGN_RF = TRUE,
  rebuild_PGN_XGB = TRUE,
  relearn_pg_embedding = TRUE,
  save_file = TRUE,
 pg_method = c("Lasso", "RF", "XGBoost"),
 python_env = "scPOEM_env"
)
```

## Arguments

mode

The mode indicating whether to analyze data from a single condition or to compare two conditions.

input\_data

A list of input data.

If mode = "single", input\_data must be a list containing the following **seven objects**:

- X: The scATAC-seq data, sparse matrix.
- Y: The scRNA-seq data, sparse matrix.
- peak\_data: A data.frame containing peak information.

14 scPOEM

• gene\_data: A data.frame containing gene information (must contain a column "gene\_name").

- cell\_data: A data.frame containing cell metadata.
- neibor\_peak: The peak IDs within a certain range of each gene, must have cols c("gene\_name", "start\_use", "end\_use"). The id numbers in "start\_use" and "end\_use" are start from 0.
- genome: The genome length for the species.

If mode = "compare", input\_data must be a **named list of two elements**, with names corresponding to two state names (e.g., "S1" and "S2"). Each element must itself be a list containing the same seven components as described above for mode = "single".

dirpath The folder path to read or write file.

count\_device The number of cpus used to train models.

nComp The number of PCs used for regression in constructing GGN.

seed An integer specifying the random seed to ensure reproducible results.

numwalks Number of random walks per node. Default is 5.

walklength Length of walk depth. Default is 3.

epochs Number of training epochs. Default is 100.

neg\_sample Number of negative samples per positive sample. Default is 5.

batch\_size Batch size for training. Default is 32.

weighted Whether the sampling network is weighted. Default is TRUE.

exclude\_pos Whether to exclude positive samples from negative sampling. Default is FALSE.

d The dimension of latent space. Default is 100.

rebuild\_GGN Logical. Whether to rebuild the gene-gene network from scratch. If FALSE,

the function will attempt to read from GGN.mtx under dirpath/test in single

mode or dirpath/state\_name/test in compare mode.

rebuild\_PPN Logical. Whether to rebuild the peak-peak network from scratch. If FALSE,

the function will attempt to read from PPN.  $\mathtt{mtx}$  under  $\mathtt{dirpath/test}$  in  $\mathtt{single}$ 

mode or dirpath/state\_name/test in compare mode.

rebuild\_PGN\_Lasso

Logical. Whether to rebuild the peak-gene network via Lasso from scratch. If FALSE, the function will attempt to read from PGN\_Lasso.mtx under

dirpath/test in single mode or dirpath/state\_name/test in compare mode.

rebuild\_PGN\_RF Logical. Whether to rebuild the peak-gene network via random forest from

scratch. If FALSE, the function will attempt to read from PGN\_RF.mtx under dirpath/test in single mode or dirpath/state\_name/test in compare

mode.

rebuild\_PGN\_XGB

Logical. Whether to rebuild the peak-gene network via XGBoost from scratch. If FALSE, the function will attempt to read from PGN\_XGB.mtx under

 ${\tt dirpath/test}\ in\ single\ mode\ or\ {\tt dirpath/state\_name/test}\ in\ compare\ mode.$ 

relearn\_pg\_embedding

Logical. Whether to relearn the low-dimensional representations for peaks and genes from scratch. If FALSE, the function will attempt to read from node\_embeddings.mtx, node\_used\_peak.csv, node\_used\_gene.csv under dirpath/embedding in single mode or

dirpath/state\_name/embedding in compare mode.

scPOEM 15

save\_file Logical, whether to save the output to a file.

pg\_method The vector of methods used to construct peak-gene net. Default is c("Lasso", "RF", "XGBoost").

python\_env Name or path of the Python environment to be used.

#### Value

The scPOEM result.

**Single Mode** Returns a list containing the following elements:

E Low-dimensional representations of peaks and genes.

peak\_node Peak IDs that are associated with other peaks or genes.

gene\_node Gene IDs that are associated with other peaks or genes.

**Compare Mode** Returns a list containing the following elements:

state1 name The single-mode result for the first condition. state2 name The single-mode result for the second condition. compare A summary list containing:

E\_g2 Low-dimensional embedding representations of genes under the two conditions. common\_genes Genes shared between both conditions and used in the analysis. diffRegulation A list of differential regulatory information for each gene.

```
library(scPOEM)
library(monocle)
dirpath <- "./example_data"</pre>
# An example for analysing a single dataset.
# Download and read data.
data(example_data_single)
single_result <- scPOEM(mode = "single",</pre>
                         input_data=example_data_single,
                         dirpath=file.path(dirpath, "single"),
                         save_file=FALSE)
# An example for analysing and comparing datasets from two conditions.
# Download compare mode example data
data(example_data_compare)
compare_result <- scPOEM(mode = "compare",</pre>
                          input_data=example_data_compare,
                          dirpath=file.path(dirpath, "compare"),
                          save_file=FALSE)
```

# **Index**

```
* datasets
example_data_compare, 4
example_data_single, 5

align_embedding, 2

eNN, 4
example_data_compare, 4
example_data_single, 5

GGN, 5

pg_embedding, 10
PGN_Lasso, 6
PGN_RF, 7
PGN_XGBoost, 9
PPN, 11

scPOEM, 13
```