

# Package ‘ProFAST’

January 11, 2024

**Type** Package

**Title** Probabilistic Factor Analysis for Spatially-Aware Dimension Reduction

**Version** 1.3

**Date** 2024-01-11

**Author** Wei Liu [aut, cre],  
Xiao Zhang [aut],  
Jin Liu [aut]

**Maintainer** Wei Liu <wei.liu@duke-nus.edu.sg>

**Description** Probabilistic factor analysis for spatially-aware dimension reduction across multi-section spatial transcriptomics data with millions of spatial locations.  
More details can be referred to Wei Liu, et al. (2023) <[doi:10.1101/2023.07.11.548486](https://doi.org/10.1101/2023.07.11.548486)>.

**License** GPL-3

**Depends** R (>= 4.0.0), gtools

**Imports** Rcpp (>= 1.0.10), AnnotationDbi, DR.SC, Matrix, mclust, PRECAST, pbapply, biomaRt, irlba, org.Hs.eg.db, org.Mm.eg.db, Seurat, parallel, performance, harmony, nnet, methods, stats, utils

**URL** <https://github.com/feiyong/ProFAST>

**BugReports** <https://github.com/feiyong/ProFAST/issues>

**Suggests** knitr, rmarkdown

**LinkingTo** Rcpp, RcppArmadillo

**VignetteBuilder** knitr

**Encoding** UTF-8

**RoxygenNote** 7.1.2

**NeedsCompilation** yes

**Repository** CRAN

**Date/Publication** 2024-01-11 11:30:18 UTC

## R topics documented:

AddParSettingFAST	2
FAST	3
FAST_run	3
FAST_structure	5
get_r2_mcfadden	6
IntegrateSRTData	7
iscmeb_run	8
model_set_FAST	9
RunHarmonyLouvain	10
RuniSCMEB	11
SelectHKgenes	11
transferGeneNames	12

<b>Index</b>	<b>13</b>
--------------	-----------

---

AddParSettingFAST	<i>Add FAST model settings for a PRECASTObj object</i>
-------------------	--

---

### Description

Add FAST model settings for a PRECASTObj object

### Usage

```
AddParSettingFAST(PRECASTObj, ...)
```

### Arguments

PRECASTObj	a PRECASTObj object created by <a href="#">CreatePRECASTObject</a> .
...	other arguments to be passed to <a href="#">model_set_FAST</a> function.

### Value

Return a revised PRECASTObj object with slot parameterList changed.

### References

None

---

FAST	<i>Run FAST model for a PRECASTObj object</i>
------	---

---

**Description**

Run FAST model for a PRECASTObj object

**Usage**

```
FAST(PRECASTObj, q = 15, fit.model = c("poisson", "gaussian"))
```

**Arguments**

PRECASTObj	a PRECASTObj object created by <a href="#">CreatePRECASTObject</a> .
q	an optional integer, specify the number of low-dimensional embeddings to extract in FAST
fit.model	an optional string, specify the version of FAST to be fitted. The Gaussian version models the log-count matrices while the Poisson version models the count matrices; default as poisson.

**Value**

Return a revised PRECASTObj object with slot PRECASTObj@resList added by a FAST component.

**References**

None

---

FAST_run	<i>(Variational) ICM-EM algorithm for implementing FAST model</i>
----------	---

---

**Description**

(Variational) ICM-EM algorithm for implementing FAST model

**Usage**

```
FAST_run(
  XList,
  AdjList,
  q = 15,
  fit.model = c("gaussian", "poisson"),
  AList = NULL,
  maxIter = 25,
```

```

    epsLogLik = 1e-05,
    verbose = TRUE,
    seed = 1,
    error_heter = TRUE,
    Psi_diag = FALSE,
    Vint_zero = FALSE
  )

```

### Arguments

<code>XList</code>	an M-length list consisting of multiple matrices with class <code>dgCMatrx</code> or <code>matrix</code> that specifies the count/log-count gene expression matrix for each data batch used for FAST model.
<code>AdjList</code>	an M-length list of sparse matrices with class <code>dgCMatrx</code> , specify the adjacency matrix used for intrinsic CAR model in FAST. We provide this interface for those users who would like to define the adjacency matrix by themselves.
<code>q</code>	an optional integer, specify the number of low-dimensional embeddings to extract in FAST. Larger <code>q</code> means more information extracted.
<code>fit.model</code>	an optional string, specify the version of FAST to be fitted. The Gaussian version models the log-count matrices while the Poisson verions models the count matrices; default as <code>gaussian</code> due to faster computation.
<code>AList</code>	an optional list with each component being a vector whose length is equal to the rows of component in <code>XList</code> , specify the normalization factor in FAST. The default is <code>NULL</code> that means the normalization factor equal to 1.
<code>maxIter</code>	the maximum iteration of ICM-EM algorithm. The default is 30.
<code>epsLogLik</code>	an optional positive vlaue, tolerance of relative variation rate of the observed pseudo loglikelihood value, default as <code>'1e-5'</code> .
<code>verbose</code>	a logical value, whether output the information in iteration.
<code>seed</code>	a postive integer, the random seed to be set in initialization.
<code>error_heter</code>	a logical value, whether use the heterogenous error for FAST model, default as <code>TRUE</code> . If <code>error.heter=FALSE</code> , then the homogenous error is used.
<code>Psi_diag</code>	a logical value, whether set the conditional covariance matrix of the intrinsic CAR to diagonal, default as <code>FALSE</code> .
<code>Vint_zero</code>	an optional logical value, specify whether the intial value of intrinsic CAR component is set to zero; default as <code>FALSE</code> .

### Details

None

### Value

return a list including the following components: (1) `hV`: an M-length list consisting of spatial embeddings in FAST; (2) `nu`: the estimated intercept vector; (3) `Psi`: the estimated covariance matrix; (4) `W`: the estimated shared loading matrix; (5) `Lam`: the estimated covariance matrix of error term; (6) `ELBO`: the ELBO value when algorithm convergence; (7) `ELBO_seq`: the ELBO values for all itrations.

**References**

None

**See Also**[FAST\\_structure](#), [FAST](#), [model\\_set\\_FAST](#)


---

FAST_structure	<i>(Variational) ICM-EM algorithm for implementing FAST model with structured parameters</i>
----------------	--

---

**Description**

(Variational) ICM-EM algorithm for implementing FAST model with structured parameters

**Usage**

```
FAST_structure(
  XList,
  AdjList,
  q = 15,
  fit.model = c("poisson", "gaussian"),
  parameterList = NULL
)
```

**Arguments**

XList	an M-length list consisting of multiple matrices with class dgCMatrix or matrix that specify the count/log-count gene expression matrix for each data batch used for FAST model.
AdjList	an M-length list of sparse matrices with class dgCMatrix, specify the adjacency matrix used for intrinsic CAR model in FAST. We provide this interface for those users who would like to define the adjacency matrix by themselves.
q	an optional integer, specify the number of low-dimensional embeddings to extract in FAST
fit.model	an optional string, specify the version of FAST to be fitted. The Gaussian version models the log-count matrices while the Poisson version models the count matrices; default as gaussian due to faster computation.
parameterList	an optional list, specify other parameters in FAST model; see <a href="#">model_set_FAST</a> for other parameters. The default is NULL that means the default parameters produced by <a href="#">model_set_FAST</a> is used.

**Details**

None

**Value**

return a list including the following components: (1) hV: an M-length list consisting of spatial embeddings in FAST; (2) nu: the estimated intercept vector; (3) Psi: the estimated covariance matrix; (4) W: the estimated shared loading matrix; (5) Lam: the estimated covariance matrix of error term; (6) ELBO: the ELBO value when algorithm convergence; (7) ELBO\_seq: the ELBO values for all iterations.

**References**

None

**See Also**

[FAST\\_run](#), [FAST](#), [model\\_set\\_FAST](#)

---

get\_r2\_mcfadden

*Calculate the the adjusted McFadden's pseudo R-square*

---

**Description**

Calculate the the adjusted McFadden's pseudo R-square between the embeddings and the labels

**Usage**

```
get_r2_mcfadden(embeds, y)
```

**Arguments**

embeds	a n-by-q matrix, specify the embedding matrix.
y	a n-length vector, specify the labels.

**Details**

None

**Value**

return the adjusted McFadden's pseudo R-square.

**References**

McFadden, D. (1987). Regression-based specification tests for the multinomial logit model. *Journal of econometrics*, 34(1-2), 63-82.

---

IntegrateSRTData      *Integrate multiple SRT data into a Seurat object*

---

### Description

Integrate multiple SRT data based on the PRECASTObj object by FAST and other model fitting.

### Usage

```
IntegrateSRTData(
  PRECASTObj,
  seulist_HK,
  Method = c("iSC-MEB", "HarmonyLouvain"),
  seuList_raw = NULL,
  covariates_use = NULL,
  Tm = NULL,
  subsample_rate = 1,
  verbose = TRUE
)
```

### Arguments

PRECASTObj	a PRECASTObj object created by <a href="#">CreatePRECASTObject</a> .
seulist_HK	a list with Seurat object as component including only the housekeeping genes.
Method	a string, specify the method to be used and two methods are supported: iSC-MEB and HarmonyLouvain. The default is iSC-MEB.
seuList_raw	an optional list with Seurat object, the raw data.
covariates_use	a string vector, the colnames in PRECASTObj@seulist[[1]]@meta.data, representing other biological covariates to considered when removing batch effects. This is achieved by adding additional covariates for biological conditions in the regression, such as case or control. Default as 'NULL', denoting no other covariates to be considered.
Tm	an optional numeric vector with the length equal to PRECASTObj@seulist, the time point information if the data include the temporal information. Default as NULL that means there is no temporal information.
subsample_rate	a real ranging in (0,1], specify the rate of spot drawing for speeding up the computation when the number of spots is very large. Default is 1, meaning using all spots.
verbose	an optional logical value, default as TRUE.

### Details

If seuList\_raw is not equal NULL or PRECASTObj@seuList is not NULL, this function will remove the unwanted variations for all genes in seuList\_raw object. Otherwise, only the the unwanted variation of genes in PRECASTObj@seulist will be removed. The former requires a big memory to

be run, while the latter not. To speed up the computation when the number of spots is very large, we also provide a subsampling schema controlled by the argument `subsample_rate`. When the total number of spots is larger than 80,000, this function will automatically draw 50,000 spots to calculate the parameters in the spatial linear model for removing unwanted variations.

### Value

Return a Seurat object by integrating all SRT data batches into a SRT data, where the column "batch" in the meta.data represents the batch ID, and the column "cluster" represents the clusters. The embeddings are put in `seu@reductions` slot and `Idents(seu)` is set to cluster label. Note that only the normalized expression is valid in the data slot while count is invalid.

---

iscmeb\_run

*Fit an iSC-MEB model using specified multi-section embeddings*

---

### Description

Integrate multiple SRT data based on the PRECASTObj by ProFAST and iSC-MEB model fitting.

### Usage

```
iscmeb_run(
  VList,
  AdjList,
  K,
  beta_grid = seq(0, 5, by = 0.2),
  maxIter = 25,
  epsLogLik = 1e-05,
  verbose = TRUE,
  int.model = "EEE",
  init.start = 1,
  Sigma_equal = FALSE,
  Sigma_diag = TRUE,
  seed = 1
)
```

### Arguments

- |         |   |
|---------|---|
| VList   | a M-length list of embeddings. The i-th element is a $n_i \times q$ matrix, where $n_i$ is the number of spots of sample i, and q is the number of embeddings. We provide this interface for those users who would like to define the embeddings by themselves. |
| AdjList | an M-length list of sparse matrices with class <code>dgCMatrix</code> , specify the adjacency matrix used for intrinsic CAR model in ProFAST. We provide this interface for those users who would like to define the adjacency matrix by themselves.            |
| K       | an integer, specify the number of clusters.   |



beta_grid	an optional vector of positive value, the candidate set of the smoothing parameter to be searched by the grid-search optimization approach, default as a sequence starts from 0, ends with 5, increase by 0.2.
maxIter	the maximum iteration of ICM-EM algorithm. The default is 25.
epsLogLik	a string, the species, one of 'Human' and 'Mouse'.
verbose	an optional integer, specify the number of housekeeping genes to be selected.
int.model	an optional string, specify which Gaussian mixture model is used in evaluating the initial values for iSC.MEB, default as "EEE"; and see <a href="#">Mclust</a> for more models' names.
init.start	an optional number of times to calculate the initial value (1 by default). When init.start is larger than 1, initial value will be determined by log likelihood of mclust results.
Sigma_equal	an optional logical value, specify whether Sigmaks are equal, default as FALSE.
Sigma_diag	an optional logical value, specify whether Sigmaks are diagonal matrices, default as TRUE.
seed	an optional integer, the random seed in fitting iSC-MEB model.

**Value**

returns a iSCMEBResObj object which contains all model results.

---

model_set_FAST	<i>Set parameters for FAST model</i>
----------------	--------------------------------------

---

**Description**

Prepare parameters setup for FAST model fitting.

**Usage**

```
model_set_FAST(
  maxIter = 30,
  epsLogLik = 1e-05,
  error_heter = TRUE,
  Psi_diag = FALSE,
  verbose = TRUE,
  seed = 1
)
```

**Arguments**

maxIter	the maximum iteration of ICM-EM algorithm. The default is 30.
epsLogLik	an optional positive value, tolerance of relative variation rate of the observed pseudo loglikelihood value, default as '1e-5'.

error_heter	a logical value, whether use the heterogenous error for FAST model, default as TRUE. If error.heter=FALSE, then the homogenous error is used.
Psi_diag	a logical value, whether set the conditional covariance matrices of intrinsic CAR to diagonal, default as FALSE
verbose	a logical value, whether output the information in iteration.
seed	a positive integer, the random seed to be set in initialization.

**Value**

return a list including the parameters set in the arguments.

**Examples**

```
model_set_FAST(maxIter = 30, epsLogLik = 1e-5,
  error_heter=TRUE, Psi_diag=FALSE, verbose=TRUE, seed=2023)
```

---

RunHarmonyLouvain	<i>Embedding alignment and clustering based on the embeddings from ProFAST</i>
-------------------	--

---

**Description**

Embedding alignment and clustering using the Harmony and Louvain based on the embeddings from ProFAST as well as determining the number of clusters.

**Usage**

```
RunHarmonyLouvain(PRECASTObj, resolution = 0.5)
```

**Arguments**

PRECASTObj	a PRECASTObj object created by <a href="#">CreatePRECASTObject</a> .
resolution	an optional real, the value of the resolution parameter, use a value above (below) 1.0 if you want to obtain a larger (smaller) number of communities.

**Value**

Return a revised PRECASTObj object with slot PRECASTObj@resList added by a Harmony component (including the aligned embeddings and embeddings of batch effects) and a Louvain component (including the clusters).

---

 RuniSCMEB

*Fit an iSC-MEB model using the embeddings from ProFAST*


---

**Description**

Fit an iSC-MEB model using the embeddings from ProFAST and the number of clusters obtained by Louvain.

**Usage**

```
RuniSCMEB(PRECASTObj, ...)
```

**Arguments**

PRECASTObj      a PRECASTObj object created by [CreatePRECASTObject](#).  
 ...              other arguments passed to [iscmeb\\_run](#).

**Value**

Return a revised PRECASTObj object with an added component iSCMEB in the slot PRECASTObj@resList (including the aligned embeddings, clusters and posterior probability matrix of clusters).

---

 SelectHKgenes

*Select housekeeping genes*


---

**Description**

Select housekeeping genes for preparation of removing unwanted variations in expression matrices

**Usage**

```
SelectHKgenes(seuList, species = c("Human", "Mouse"), HK.number = 200)
```

**Arguments**

seuList            an M-length list consisting of Seurat object, include the information of expression matrix and spatial coordinates (named row and col) in the slot meta.data.  
 species            a string, the species, one of 'Human' and 'Mouse'.  
 HK.number         an optional integer, specify the number of housekeeping genes to be selected.

**Value**

Return a string vector of the selected gene names.

---

transferGeneNames      *Transfer gene names from one format to the other format*

---

### Description

Transfer gene names from one format to the other format for two species: human and mouse.

### Usage

```
transferGeneNames(  
  genelist,  
  now_name = "ensembl",  
  to_name = "symbol",  
  species = c("Human", "Mouse"),  
  Method = c("eg.db", "biomart")  
)
```

### Arguments

genelist	a string vector, the gene list to be transferred.
now_name	a string, the current format of gene names, one of 'ensembl', 'symbol'.
to_name	a string, the format of gene names to transfer, one of 'ensembl', 'symbol'.
species	a string, the species, one of 'Human' and 'Mouse'.
Method	a string, the method to use, one of 'biomaRt' and 'eg.db', default as 'eg.db'.

### Value

Return a string vector of transferred gene names. The gene names not matched in the database will not change.

### Examples

```
geneNames <- c("ENSG00000171885", "ENSG00000115756")  
transferGeneNames(geneNames, now_name = "ensembl", to_name="symbol", species="Human", Method='eg.db')
```

# Index

AddParSettingFAST, [2](#)

CreatePRECASTObject, [2](#), [3](#), [7](#), [10](#), [11](#)

FAST, [3](#), [5](#), [6](#)

FAST\_run, [3](#), [6](#)

FAST\_structure, [5](#), [5](#)

get\_r2\_mcfadden, [6](#)

IntegrateSRTData, [7](#)

iscmeb\_run, [8](#), [11](#)

Mclust, [9](#)

model\_set\_FAST, [2](#), [5](#), [6](#), [9](#)

RunHarmonyLouvain, [10](#)

RuniSCMEB, [11](#)

SelectHKgenes, [11](#)

transferGeneNames, [12](#)