

# Package ‘rDppDiversity’

June 1, 2021

**Title** Subset Searching Algorithm Using DPP Greedy MAP

**Version** 0.0.2

**Description** Given item set, item representation vector, and item ratings, find a subset with better relevance-diversity trade-off. Also provide machine learning algorithm to learn item representations maximizing log likelihood under DPP assumption.

References: [1]Laming Chen, Guoxin Zhang, and Hanning Zhou(2017)<[https://lrsr2017.files.wordpress.com/2017/08/lrsr\\_2017\\_lamingchen.pdf](https://lrsr2017.files.wordpress.com/2017/08/lrsr_2017_lamingchen.pdf)>

[2]Laming Chen, Guoxin Zhang, and Hanning Zhou(2018)<<https://papers.nips.cc/paper/2018/file/dbbf603ff0e99629dda5d75b6f75f966-Paper.pdf>>

[3]Wilhelm, Mark & Ramanathan, Ajith & Bonomo, Alexander & Jain, Sagar & Chi, Ed & Gillenwater, Jennifer(2018)<[doi:10.1145/3269206.3272018](https://doi.org/10.1145/3269206.3272018)>.

**License** MIT + file LICENSE

**Encoding** UTF-8

**LazyData** false

**RoxygenNote** 7.1.1

**LinkingTo** Rcpp

**Imports** Rcpp

**NeedsCompilation** yes

**Author** Sining Ng [aut, cre] (<<https://orcid.org/0000-0003-3463-707X>>),

Laming Chen [aut],

Guoxin Zhang [aut],

Hanning Zhou [aut],

Jennifer Gillenwater [aut],

Ed H. Chi [aut],

Sagar Jain [aut],

Alexander Bonomo [aut],

Ajith Ramanathan [aut],

Mark Wilhelm [aut],

Antonio Sánchez [cph] (<https://gitlab.com/cantonios>),

Christoph Hertzberg [cph] (<https://gitlab.com/chhtz>),

David Tellenbach [cph] (<https://gitlab.com/tellenbach>),

Eigen Bugzilla [cph] (<https://gitlab.com/eigenbz>),

Gael Guennebaud [cph] (<https://gitlab.com/ggael>),  
 Rasmus Munk Larsen [cph] (<https://gitlab.com/rmlarsen1>),  
 Guoqiang QI [cph] (<https://gitlab.com/guoqiangqi1>),  
 Konstantinos Margaritis [cph] (<https://gitlab.com/freevec>)

**Maintainer** Sining Ng <sining\_ng@icloud.com>

**Repository** CRAN

**Date/Publication** 2021-06-01 10:50:02 UTC

## R topics documented:

bestSubset . . . . . 2  
 learnItemEmb . . . . . 3

**Index** . . . . . 4

---

bestSubset	<i>Given item set, item representation vector, and item ratings, find a subset with better relevance-diversity trade-off.</i>
------------	---

---

### Description

Given item set, item representation vector, and item ratings, find a subset with better relevance-diversity trade-off.

### Usage

```
bestSubset(item_representations, ratings, n)
```

### Arguments

item\_representations  
 matrix with nrow = num\_items and ncol = feature\_dim

ratings  
 matrix with nrow = num\_items and ncol = 1

n  
 int. subset size

### Value

dataframe with 2 columns: id & gain

### Examples

```
library(rDppDiversity)
bestSubset(matrix(c(-0.1, -0.15, 0.2, 0.3, -0.2, -0.3), nrow=3, ncol=2, byrow=TRUE), c(1, 2, 3), 3)
```

---

learnItemEmb	<i>Machine learning algorithm to learn item representations maximizing log likelihood under DPP assumption.</i>
--------------	---

---

### Description

Machine learning algorithm to learn item representations maximizing log likelihood under DPP assumption.

### Usage

```
learnItemEmb(  
  train_data_path,  
  emb_size,  
  regularization,  
  learning_rate,  
  neg_sample_cnt,  
  epoch  
)
```

### Arguments

train_data_path	A string for text file path. Each line: item_id,item_id,item_id
emb_size	int. ColumnNum for model parameter. While RowNum = number of uniq items parsed in train_data_path
regularization	float. Default = 0.1
learning_rate	float. Generally begin with small learning_rate will train better.
neg_sample_cnt	int.
epoch	int.

### Value

A list contains 1) learned item embedding matrix; 2) item names vector; 3) log likelihood on each training step vector.

### Examples

```
library(rDppDiversity)  
data_path=system.file("extdata", "data.txt", package = "rDppDiversity")  
learnItemEmb(data_path, 3, 0.1, 0.01, 0, 10)
```

# Index

`bestSubset`, 2

`learnItemEmb`, 3