

# Package ‘Deconfounder’

November 18, 2020

**Title** Deconfounder

**Version** 0.0.0.9000

**Description** Deconfounder is a causal inference model for estimating the treatment effects of medications with electronic health records. We aim to assess the treatment effects of multiple medications on outcomes of interest (e.g., lab measurements) using the Deconfounder. Deconfounder identifies the causal medications that have either direct effect or adverse effect on each clinical measurement. The inputs to Deconfounder are medication records and pre-treatment and post-treatment measurement values. Deconfounder fits a probabilistic factor model (e.g., poisson factorization or deep exponential family) to the medication records to construct substitute confounders and adjusts for substitute confounders in the outcome model for assessing the causal effects of medications..

**License** Apache License 2.0

**RoxygenNote** 7.0.0

## R topics documented:

createCohorts . . . . .	1
fitDeconfounder . . . . .	3
generateData . . . . .	4
preprocessingData . . . . .	5

<b>Index</b>	<b>7</b>
--------------	----------

---

createCohorts	<i>Generate cohort</i>
---------------	------------------------

---

## Description

Generate cohort

## Usage

```
createCohorts(  
  connection,  
  cdmDatabaseSchema,  
  oracleTempSchema = NULL,  
  vocabularyDatabaseSchema = cdmDatabaseSchema,  
  cohortDatabaseSchema,
```

```

    targetCohortTable,
    createTargetCohortTable = TRUE,
    conditionConceptIds,
    measurementConceptId,
    observationWindowBefore,
    observationWindowAfter,
    drugWindow,
    targetCohortId
  )

```

### Arguments

connection	Connection to the database. An object of class connect as created by the connect function in the DatabaseConnector package.
cdmDatabaseSchema	A schema where OMOP CDM data are stored.
oracleTempSchema	A schema that can be used to create temp tables in when using Oracle.
vocabularyDatabaseSchema	A schema where vocabulary is stored
cohortDatabaseSchema	A schema where the cohort is stored
targetCohortTable	A string corresponds to the name of cohort table
createTargetCohortTable	A boolean that indicates whether the targetCohortTable will be created. Default TRUE.
conditionConceptIds	A list of condition concept IDs that correspond to the disease of interest.
measurementConceptId	A numeric of measurement concept ID
observationWindowBefore	An integer indicates the number of pre-treatment days used to look for pre-treatment measurement.
observationWindowAfter	An integer indicates the number of post-treatment days used to look for post-treatment measurement.
drugWindow	An integer indicates the number of post-treatment days during which drug exposure are also considered. Default is 0
targetCohortId	An integer of the cohort ID.

### Details

This function generates cohort for the medical deconfounder (single outcome)

### Value

Returns a string containing the rendered SQL.

---

fitDeconfounder	<i>fit the deconfounder to estimate average treatment effect</i>
-----------------	--

---

## Description

fit the deconfounder to estimate average treatment effect

## Usage

```
fitDeconfounder(
  data_dir,
  save_dir,
  factor_model,
  learning_rate = 1e-04,
  max_steps = 1e+05,
  latent_dim = 1,
  layer_dim = c(30, 10),
  batch_size = 1024,
  num_samples = 1,
  holdout_portion = 0.5,
  print_steps = 50,
  tolerance = 3,
  num_confounder_samples = 30,
  CV = 5,
  outcome_type = "linear"
)
```

## Arguments

data_dir	String: the directory where cohort data are stored
save_dir	String: the directory where results will be stored
factor_model	String: the type of probabilistic factor model to fit. Choices are: PMF or DEF.
learning_rate	Float: The learning rate for the probabilistic factor model.
max_steps	Integer: the maximum steps to run the probabilistic factor model.
latent_dim	Integer: the number of latent dimensions in PMF.
layer_dim	List: a list of length 2. The number of latent dimensions in each layer of the 2-layer DEF.
batch_size	Integer: the number of datapoints to use in each training step of the probabilistic model.
num_samples	Integer: number of samples from variational distribution used in updating variational parameters.
holdout_portion	Float: A value between 0 and 1. The proportion of data heldout for predictive model checking in checking the probabilistic model.
print_steps	Integer: Print the results during training.
tolerance	Integer: The termination criteria for training the probabilistic model.

num_confounder_samples	Integer: number of samples of substitute confounder from the posterior, input for the outcome model for estimating ATE.
CV	Integer: Fold of cross validation in the outcome model.
outcome_type	String: The type of outcome. Choices are: linear

---

generateData

*Generate data*


---

## Description

Generate data

## Usage

```
generateData(
  connection,
  cdmDatabaseSchema,
  oracleTempSchema = NULL,
  vocabularyDatabaseSchema = cdmDatabaseSchema,
  cohortDatabaseSchema,
  targetCohortTable,
  drugExposureTable,
  measurementTable,
  conditionConceptIds,
  measurementConceptId,
  observationWindowBefore,
  observationWindowAfter,
  drugWindow,
  createTargetCohortTable = T,
  createTargetCohort = T,
  extractFeature = T,
  targetCohortId = NULL,
  dataFolder,
  drugFilename,
  measFilename
)
```

## Arguments

connection	Connection to the database. An object of class connect as created by the connect function in the DatabaseConnector package.
cdmDatabaseSchema	A schema where OMOP CDM data are stored.
oracleTempSchema	A schema that can be used to create temp tables in when using Oracle.
vocabularyDatabaseSchema	A schema where vocabulary is stored
cohortDatabaseSchema	A schema where the cohort is stored

targetCohortTable	A string corresponds to the name of the cohort table
drugExposureTable	A string corresponds to the name of the drug exposure table
measurementTable	A string corresponds to the name of the measurement table
conditionConceptIds	A list of condition concept IDs that correspond to the disease of interest.
measurementConceptId	A numeric of measurement concept ID
observationWindowBefore	An integer indicates the number of pre-treatment days used to look for pre-treatment measurement.
observationWindowAfter	An integer indicates the number of post-treatment days used to look for post-treatment measurement.
drugWindow	An integer indicates the number of post-treatment days during which drug exposure are also considered. Default is 0
createTargetCohortTable	A boolean that indicates whether the targetCohortTable will be created. Default TRUE.
createTargetCohort	A boolean that indicates whether to create the cohort. Default TRUE.
extractFeature	A boolean that indicates whether to extract features for the cohort. Default TRUE.
targetCohortId	An integer of the cohort ID.
dataFolder	A string indicates where output will be stored.
drugFilename	A string for the name of the table to store drug exposure.
measFilename	A string for the name of the table to store measurement.

## Details

This function generates cohort and extracts data of the cohort.

## Value

Returns a string containing the rendered SQL.

---

preprocessingData	<i>Preprocess data</i>
-------------------	------------------------

---

## Description

Preprocess data

## Usage

```
preprocessingData(dataFolder, measFilename, drugFilename, drugWindow)
```

**Arguments**

<code>dataFolder</code>	A string indicates where output will be stored.
<code>measFilename</code>	A string for the name of the table to store measurement.
<code>drugFilename</code>	A string for the name of the table to store drug exposure.
<code>drugWindow</code>	An integer indicates the number of post-treatment days during which drug exposure are also considered. Default is 0

# Index

`createCohorts`, [1](#)  
`fitDeconfounder`, [3](#)  
`generateData`, [4](#)  
`preprocessingData`, [5](#)