

Package ‘PheValuator’

December 1, 2022

Type Package

Title Rapid Phenotype Evaluation

Version 2.1.13

Author Joel N. Swerdel

Maintainer Joel N. Swerdel <jswerdel@its.jnj.com>

Description A package for evaluating phenotype algorithms. The package develops a diagnostic prediction model using a set of noisy positives and noisy negatives. It can then apply the model to a large cohort to create a set of subjects each with a predicted probability for the health outcome of interest. This cohort can then be used to evaluate phenotype algorithms through the calculation of the necessary performance characteristics - sensitivity, specificity, and positive and negative predictive value.

License Apache License 2.0

VignetteBuilder knitr

Depends R (>= 3.2.2),
DatabaseConnector,
FeatureExtraction (>= 2.1.5)

Imports PatientLevelPrediction (>= 6.0.4),
SqlRender (>= 1.5.2),
dplyr (>= 0.7.8),
methods,
ParallelLogger,
tibble

Suggests knitr,
rmarkdown,
ROhdsiWebApi,
shiny,
shinydashboard,
DT,
htmltools,
qpdf,
testthat,
Eunomia,
CohortGenerator,
withr

Remotes ohdsi/PatientLevelPrediction,
 ohdsi/FeatureExtraction,
 ohdsi/ROhdsiWebApi,
 ohdsi/Eunomia,
 ohdsi/CohortGenerator

URL <https://ohdsi.github.io/PheValuator>, <https://github.com/OHDSI/PheValuator>

BugReports <https://github.com/OHDSI/PheValuator/issues>

RoxygenNote 7.2.2

Encoding UTF-8

R topics documented:

computePerformanceMetricsFromCounts	2
createCreateEvaluationCohortArgs	3
createDefaultCovariateSettings	5
createEvaluationCohort	6
createPheValuatorAnalysis	8
createTestPhenotypeAlgorithmArgs	9
loadPheValuatorAnalysisList	10
runPheValuatorAnalyses	10
savePheValuatorAnalysisList	11
summarizePheValuatorAnalyses	12
testPhenotypeAlgorithm	12

Index	14
--------------	-----------

computePerformanceMetricsFromCounts

Compute performance metrics based on a 2-by-2 counts table

Description

Compute performance metrics based on a 2-by-2 counts table

Usage

```
computePerformanceMetricsFromCounts(counts)
```

Arguments

counts Counts as created by the `testPhenotypeAlgorithm` function.

Value

A tibble with the statistics metrics added to the counts table.

```
createCreateEvaluationCohortArgs
```

Create a parameter object for the function createEvaluationCohort

Description

Create a parameter object for the function createEvaluationCohort

Usage

```
createCreateEvaluationCohortArgs(
  xSpecCohortId,
  xSensCohortId,
  prevalenceCohortId,
  xSpecCohortSize = 5000,
  covariateSettings = createDefaultCovariateSettings(excludedCovariateConceptIds = c(),
    addDescendantsToExclude = TRUE),
  modelPopulationCohortId = 0,
  modelPopulationCohortIdStartDay = 0,
  modelPopulationCohortIdEndDay = 0,
  evaluationPopulationCohortId = 0,
  evaluationPopulationCohortIdStartDay = 0,
  evaluationPopulationCohortIdEndDay = 0,
  modelBaseSampleSize = 15000,
  baseSampleSize = 2e+06,
  lowerAgeLimit = 0,
  upperAgeLimit = 120,
  visitLength = 0,
  visitType = c(9201, 9202, 9203, 581477, 262),
  gender = c(8507, 8532),
  race = 0,
  ethnicity = 0,
  startDate = "19001010",
  endDate = "21000101",
  modelId = "main",
  evaluationCohortId = "main",
  excludeModelFromEvaluation = FALSE,
  removeSubjectsWithFutureDates = TRUE,
  saveEvaluationCohortPlpData = FALSE
)
```

Arguments

<code>xSpecCohortId</code>	The number of the "extremely specific (xSpec)" cohortdefinition id in the cohort table (for noisy positives).
<code>xSensCohortId</code>	The number of the "extremely sensitive (xSens)" cohortdefinition id in the cohort table (for noisy negatives).
<code>prevalenceCohortId</code>	The number of the cohort definition id to determine the disease prevalence.
<code>xSpecCohortSize</code>	The recommended xSpec sample size to use in model (default = 5000)

<code>covariateSettings</code>	A covariateSettings object as generated using <code>createCovariateSettings()</code> .
<code>modelPopulationCohortId</code>	The number of the cohort to be used as a base population for the model. If set to 0, the entire database population will be used.
<code>modelPopulationCohortIdStartDay</code>	The number of days relative to the <code>mainPopulationCohortId</code> cohort start date to begin including visits.
<code>modelPopulationCohortIdEndDay</code>	The number of days relative to the <code>mainPopulationCohortId</code> cohort start date to end including visits.
<code>evaluationPopulationCohortId</code>	The number of the cohort to be used as a base population for the evaluation cohort. If set to 0, the entire database population will be used.
<code>evaluationPopulationCohortIdStartDay</code>	The number of days relative to the <code>evaluationPopulationCohortId</code> cohort start date to begin including visits.
<code>evaluationPopulationCohortIdEndDay</code>	The number of days relative to the <code>evaluationPopulationCohortId</code> cohort start date to end including visits.
<code>modelBaseSampleSize</code>	The number of non-xSpec subjects to include in the model
<code>baseSampleSize</code>	The maximum number of subjects in the evaluation cohort.
<code>lowerAgeLimit</code>	The lower age for subjects in the model.
<code>upperAgeLimit</code>	The upper age for subjects in the model.
<code>visitLength</code>	The minimum length of index visit for acute outcomes.
<code>visitType</code>	The <code>concept_id</code> for the visit type.
<code>gender</code>	The gender(s) to be included.
<code>race</code>	The race(s) to be included.
<code>ethnicity</code>	The ethnicity(s) to be included.
<code>startDate</code>	The starting date for including subjects in the model.
<code>endDate</code>	The ending date for including subjects in the model.
<code>modelId</code>	A string used to generate the file names for this model.
<code>evaluationCohortId</code>	A string used to generate the file names for this evaluation cohort.
<code>excludeModelFromEvaluation</code>	Should subjects used in the model be excluded from the evaluation cohort?
<code>removeSubjectsWithFutureDates</code>	For buggy data with data in the future: ignore subjects with dates in the future?
<code>saveEvaluationCohortPlpData</code>	Should the large PLP file for the evaluation cohort be saved? To be used for debugging purposes.

Details

Create an object defining the parameter values.

```
createDefaultCovariateSettings
```

Create the default covariate settings

Description

Create the default covariate settings for acute or chronic diseases

Usage

```
createDefaultCovariateSettings(
  excludedCovariateConceptIds = c(),
  includedCovariateIds = c(),
  includedCovariateConceptIds = c(),
  addDescendantsToExclude = FALSE,
  startDayWindow1 = 0,
  endDayWindow1 = 9999,
  startDayWindow2 = NULL,
  endDayWindow2 = NULL,
  startDayWindow3 = NULL,
  endDayWindow3 = NULL
)
```

Arguments

<code>excludedCovariateConceptIds</code>	A list of conceptIds to exclude from featureExtraction. These should include all concept_ids that were used to define the xSpec model (default=NULL)
<code>includedCovariateIds</code>	A list of covariate IDs that should be restricted to.
<code>includedCovariateConceptIds</code>	A list of covariate concept IDs that should be restricted to.
<code>addDescendantsToExclude</code>	Should descendants of excluded concepts also be excluded? (default=FALSE)
<code>startDayWindow1</code>	The day to start time window 1 for feature extraction
<code>endDayWindow1</code>	The day to end time window 1 for feature extraction
<code>startDayWindow2</code>	The day to start time window 2 for feature extraction
<code>endDayWindow2</code>	The day to end time window 2 for feature extraction
<code>startDayWindow3</code>	The day to start time window 3 for feature extraction
<code>endDayWindow3</code>	The day to end time window 3 for feature extraction #'

Details

Function to create the default covariate settings for acute or chronic diseases

```
createEvaluationCohort
```

Create the evaluation cohort

Description

Create the evaluation cohort

Usage

```
createEvaluationCohort(
  connectionDetails,
  oracleTempSchema = NULL,
  tempEmulationSchema = getOption("sqlRenderTempEmulationSchema"),
  xSpecCohortId,
  xSensCohortId,
  prevalenceCohortId,
  xSpecCohortSize = 5000,
  cdmDatabaseSchema,
  cohortDatabaseSchema,
  cohortTable,
  workDatabaseSchema,
  covariateSettings = createDefaultCovariateSettings(excludedCovariateConceptIds = c(),
    addDescendantsToExclude = TRUE),
  modelPopulationCohortId = 0,
  modelPopulationCohortIdStartDay = 0,
  modelPopulationCohortIdEndDay = 0,
  evaluationPopulationCohortId = 0,
  evaluationPopulationCohortIdStartDay = 0,
  evaluationPopulationCohortIdEndDay = 0,
  modelBaseSampleSize = 15000,
  baseSampleSize = 2e+06,
  lowerAgeLimit = 0,
  upperAgeLimit = 120,
  visitLength = 0,
  visitType = c(9201, 9202, 9203, 262, 581477),
  gender = c(8507, 8532),
  race = 0,
  ethnicity = 0,
  startDate = "19001010",
  endDate = "21000101",
  cdmVersion = "5",
  outFolder = getwd(),
  modelId = "main",
  evaluationCohortId = "main",
  excludeModelFromEvaluation = FALSE,
  removeSubjectsWithFutureDates = TRUE,
  saveEvaluationCohortPlpData = FALSE
)
```

Arguments

connectionDetails	connectionDetails created using the function createConnectionDetails in the DatabaseConnector package.
oracleTempSchema	DEPRECATED: use tempEmulationSchema instead.
tempEmulationSchema	Some database platforms like Oracle and Impala do not truly support temp tables. To emulate temp tables, provide a schema with write privileges where temp tables can be created.
xSpecCohortId	The number of the "extremely specific (xSpec)" cohort definition id in the cohort table (for noisy positives).
xSensCohortId	The number of the "extremely sensitive (xSens)" cohort definition id in the cohort table (for noisy negatives).
prevalenceCohortId	The number of the cohort definition id to determine the disease prevalence.
xSpecCohortSize	The recommended xSpec sample size to use in model (default = NULL)
cdmDatabaseSchema	The name of the database schema that contains the OMOP CDM instance. Requires read permissions to this database. On SQL Server, this should specify both the database and the schema, so for example 'cdm_instance.dbo'.
cohortDatabaseSchema	The name of the database schema that is the location where the cohort data used to define the at risk cohort is available. Requires read permissions to this database.
cohortTable	The tablename that contains the at risk cohort. The expectation is cohortTable has format of COHORT table: cohort_concept_id, SUBJECT_ID, COHORT_START_DATE, COHORT_END_DATE.
workDatabaseSchema	The name of the database schema that is the location where a table can be created and afterwards removed. Requires write permissions to this database.
covariateSettings	A covariateSettings object as generated using createCovariateSettings().
modelPopulationCohortId	The number of the cohort to be used as a base population for the model. If set to 0, the entire database population will be used.
modelPopulationCohortIdStartDay	The number of days relative to the mainPopulationCohortId cohort start date to begin including visits.
modelPopulationCohortIdEndDay	The number of days relative to the mainPopulationCohortId cohort start date to end including visits.
evaluationPopulationCohortId	The number of the cohort to be used as a base population for the evaluation cohort. If set to 0, the entire database population will be used.
evaluationPopulationCohortIdStartDay	The number of days relative to the evaluationPopulationCohortId cohort start date to begin including visits.

evaluationPopulationCohortIdEndDay	The number of days relative to the evaluationPopulationCohortId cohort start date to end including visits.
modelBaseSampleSize	The number of non-xSpec subjects to include in the model
baseSampleSize	The maximum number of subjects in the evaluation cohort.
lowerAgeLimit	The lower age for subjects in the model.
upperAgeLimit	The upper age for subjects in the model.
visitLength	The minimum length of index visit for acute outcomes.
visitType	The concept_id for the visit type.
gender	The gender(s) to be included.
race	The race(s) to be included.
ethnicity	The ethnicity(s) to be included.
startDate	The starting date for including subjects in the model.
endDate	The ending date for including subjects in the model.
cdmVersion	The CDM version of the database.
outFolder	The folder where the output files will be written.
modelId	A string used to generate the file names for this model.
evaluationCohortId	A string used to generate the file names for this evaluation cohort.
excludeModelFromEvaluation	Should subjects used in the model be excluded from the evaluation cohort?
removeSubjectsWithFutureDates	For buggy data with data in the future: ignore subjects with dates in the future?
saveEvaluationCohortPlpData	Should the large PLP file for the evaluation cohort be saved? To be used for debugging purposes.

Details

Fits a diagnostic prediction model, and uses it to create an evaluation cohort with probabilities for the health outcome of interest.

createPheValuatorAnalysis

Create a PheValuator analysis specification

Description

Create a PheValuator analysis specification

Usage

```
createPheValuatorAnalysis(
  analysisId,
  description,
  createEvaluationCohortArgs,
  testPhenotypeAlgorithmArgs
)
```


Arguments

analysisId	An integer that will be used later to refer to this specific set of analysis choices.
description	A short description of the analysis.
createEvaluationCohortArgs	An object representing the arguments to be used when calling the createCreateEvaluationCohort function.
testPhenotypeAlgorithmArgs	Should the createTestPhenotypeAlgorithmArgs function be used in this analysis?

Details

Create a set of analysis choices, to be used with the [runPheValuatorAnalyses](#) function.

```
createTestPhenotypeAlgorithmArgs
```

Create a parameter object for the function testPhenotypeAlgorithm

Description

Create a parameter object for the function testPhenotypeAlgorithm

Usage

```
createTestPhenotypeAlgorithmArgs(
  cutPoints = c("EV"),
  phenotypeCohortId,
  washoutPeriod = 0,
  splayPrior = 7,
  splayPost = 7
)
```

Arguments

cutPoints	A list of threshold predictions for the evaluations. Include "EV" for the expected value
phenotypeCohortId	The ID of the cohort to evaluate in the specified cohort table.
washoutPeriod	The minimum required continuous observation time prior to indexdate for subjects within the cohort to test (Default = 0).
splayPrior	The number of days to allow for test phenotype visit date prior to evaluation date
splayPost	The number of days to allow for test phenotype visit date after evaluation date

Details

Create an object defining the parameter values.

loadPheValuatorAnalysisList

Load a list of phevaluatorAnalysis from file

Description

Load a list of objects of type pheValuatorAnalysis from file. The file is in JSON format.

Usage

```
loadPheValuatorAnalysisList(file)
```

Arguments

file	The name of the file
------	----------------------

Value

A list of objects of type pheValuatorAnalysis.

runPheValuatorAnalyses

Run a list of analyses

Description

Run a list of analyses

Usage

```
runPheValuatorAnalyses(  
  connectionDetails,  
  oracleTempSchema = NULL,  
  tempEmulationSchema = getOption("sqlRenderTempEmulationSchema"),  
  cdmDatabaseSchema,  
  cohortDatabaseSchema = cdmDatabaseSchema,  
  cohortTable = "cohort",  
  workDatabaseSchema = cdmDatabaseSchema,  
  cdmVersion = 5,  
  outputFolder,  
  pheValuatorAnalysisList  
)
```

Arguments

connectionDetails	An R object of type connectionDetails created using the function createConnectionDetails in the DatabaseConnector package.
oracleTempSchema	DEPRECATED: use tempEmulationSchema instead.
tempEmulationSchema	Some database platforms like Oracle and Impala do not truly support temp tables. To emulate temp tables, provide a schema with write privileges where temp tables can be created.
cdmDatabaseSchema	The name of the database schema that contains the OMOP CDM instance. Requires read permissions to this database. On SQL Server, this should specify both the database and the schema, so for example 'cdm_instance.dbo'.
cohortDatabaseSchema	The name of the database schema that is the location where the cohort data used to define the at risk cohort is available. Requires read permissions to this database.
cohortTable	The tablename that contains the at risk cohort. The expectation is cohortTable has format of COHORT table: cohort_concept_id, SUBJECT_ID, COHORT_START_DATE, COHORT_END_DATE.
workDatabaseSchema	The name of the database schema that is the location where a table can be created and afterwards removed. Requires write permissions to this database.
cdmVersion	Define the OMOP CDM version used: currently supports "5".
outputFolder	Name of the folder where all the outputs will be written to.
pheValuatorAnalysisList	A list of objects of type pheValuatorAnalysis as created using the createPheValuatorAnalysis function.

Details

Run a list of analyses.

Value

A data frame specifying where the constructed evaluation cohort and phenotype evaluation results can be found in the local file system.

```
savePheValuatorAnalysisList
```

Save a list of phevaluatorAnalysis to file

Description

Write a list of objects of type pheValuatorAnalysis to file. The file is in JSON format.

Usage

```
savePheValuatorAnalysisList(pheValuatorAnalysisList, file)
```

Arguments

pheValuatorAnalysisList The pheValuatorAnalysis list to be written to file
 file The name of the file where the results will be written

summarizePheValuatorAnalyses

Summarize results of PheValuator analyses

Description

Summarize results of PheValuator analyses

Usage

```
summarizePheValuatorAnalyses(referenceTable, outputFolder)
```

Arguments

referenceTable A reference table as created using [runPheValuatorAnalyses](#).
 outputFolder The output folder used when calling [runPheValuatorAnalyses](#).

Value

A data frame of results.

testPhenotypeAlgorithm

Test phenotype algorithms

Description

Test phenotype algorithms

Usage

```
testPhenotypeAlgorithm(  
  connectionDetails,  
  cutPoints = c("EV"),  
  outFolder,  
  evaluationCohortId = "main",  
  phenotypeCohortId,  
  cdmDatabaseSchema,  
  cohortDatabaseSchema,  
  cohortTable,  
  washoutPeriod = 0,  
  splayPrior = 7,  
  splayPost = 7  
)
```

Arguments

connectionDetails	ConnectionDetails created using the function createConnectionDetails in the DatabaseConnector package.
cutPoints	A list of threshold predictions for the evaluations. Include "EV" for the expected value
outFolder	The folder where the cohort evaluation output files are written
evaluationCohortId	A string used to generate the file names for the evaluation cohort.
phenotypeCohortId	The ID of the cohort to evaluate in the specified cohort table.
cdmDatabaseSchema	The name of the database schema that contains the OMOP CDM instance. Requires read permissions to this database. On SQL Server, this should specify both the database and the schema, so for example 'cdm_instance.dbo'.
cohortDatabaseSchema	The name of the database schema that is the location where the cohort data used to define the at risk cohort is available. Requires read permissions to this database.
cohortTable	The tablename that contains the at risk cohort. The expectation is cohortTable has format of COHORT table: cohort_concept_id, SUBJECT_ID, COHORT_START_DATE, COHORT_END_DATE.
washoutPeriod	The minimum required continuous observation time prior to index date for subjects within the cohort to test (Default = 0).
splayPrior	The number of days to allow for test phenotype visit date prior to evaluation date
splayPost	The number of days to allow for test phenotype visit date after evaluation date

Details

This function will perform the phenotype algorithm evaluation using the evaluation cohort returned from createEvalCohort and the phenotype algorithm cohort specified

Value

A dataframe with the results from the phenotype algorithm evaluation.

If 0.5 is included as a cutpoint, the data frame will have an attribute called 'misses', a dataframe with a sample of subject ids for TPs, FPs, TNs, and FNs for the 50 percent and over prediction threshold.

Index

computePerformanceMetricsFromCounts, [2](#)
createCreateEvaluationCohortArgs, [3](#), [9](#)
createDefaultCovariateSettings, [5](#)
createEvaluationCohort, [6](#)
createPheValuatorAnalysis, [8](#), [11](#)
createTestPhenotypeAlgorithmArgs, [9](#), [9](#)

loadPheValuatorAnalysisList, [10](#)

runPheValuatorAnalyses, [9](#), [10](#), [12](#)

savePheValuatorAnalysisList, [11](#)
summarizePheValuatorAnalyses, [12](#)

testPhenotypeAlgorithm, [2](#), [12](#)