Package 'PheValuator'

September 27, 2019

Type Package

Title Rapid Phenotype Evaluation

Version 1.0.0

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
preditive value.

License Apache License 2.0

VignetteBuilder knitr

Encoding UTF-8

LazyData true

Depends R (>= 3.2.2)

Imports PatientLevelPrediction (>= 2.0.4),
FeatureExtraction (>= 2.1.5),
SqlRender (>= 1.5.2),
data.table (>= 1.11.8),
dplyr (>= 0.7.8),
stringr (>= 1.3.1),
DatabaseConnector (>= 2.1.4)

Suggests knitr

RoxygenNote 6.1.1

R topics documented:

```
createEvaluationCohort
```
Create the evaluation cohort

Description

Create the evaluation cohort

Usage

```
createEvaluationCohort(connectionDetails, xSpecCohort, cdmDatabaseSchema,
  cohortDatabaseSchema, cohortDatabaseTable, outDatabaseSchema,
  evaluationOutputFileName, modelOutputFileName,
  mainPopulationCohort = 0, lowerAgeLimit = 0, upperAgeLimit = 120,
  startDays = -10000, endDays = 10000, gender = c(8507, 8532),
  startDate = "19001010", endDate = "21000101", cdmVersion = "5",
  outFolder = getwd())
```

Arguments

connectionDetails

connectionDetails created using the function createConnectionDetails in the DatabaseC-onnector package.

xSpecCohort The number of the "extremely specific (xSpec)" cohort definition id in the cohort table (for noisy positives)

cdmDatabaseSchema

The name of the database schema that contains the OMOP CDM instance. Requires read permissions to this database. On SQL Server, this should specifiy 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.

cohortDatabaseTable

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.

outDatabaseSchema

The name of the database schema that is the location where the data used to define the outcome cohorts is available. Requires read permissions to this database.

evaluationOutputFileName

A string designation for the evaluation cohort file

modelOutputFileName

A string designation for the training model file

mainPopulationCohort

The number of the cohort to be used as a base population for the model (default=NULL)

lowerAgeLimit The lower age for subjects in the model (default=NULL)

upperAgeLimit The upper age for subjects in the model (default=NULL)

startDays	The days to include prior to the cohort start date (default=-10000)
endDays	The days to include after the cohort start date (default=10000)
gender	The gender(s) to be included (default c(8507, 8532))
startDate	The starting date for including subjects in the model (default=NULL)
endDate	The ending date for including subjects in the model (default=NULL)
cdmVersion	The CDM version of the database (default=5)
outFolder	The folder where the output files will be written (default=working directory)

Details

Creates the evaluation cohort and applies a diagnostic prediction model for determination of a probability for the health outcome of interest

---

createPhenotypeModel   *Create the phenotype model*

Description

Create the phenotype model

Usage

```
createPhenotypeModel(connectionDetails, xSpecCohort, cdmDatabaseSchema,
  cohortDatabaseSchema, cohortDatabaseTable, outDatabaseSchema,
  modelOutputFileName = "train", xSensCohort,
  prevalenceCohort = xSensCohort, excludedConcepts = c(),
  addDescendantsToExclude = FALSE, mainPopulationCohort = 0,
  lowerAgeLimit = 0, upperAgeLimit = 120, gender = c(8507, 8532),
  startDate = "19000101", endDate = "21000101", cdmVersion = "5",
  outFolder = getwd())
```

Arguments

connectionDetails

> connectionDetails created using the function createConnectionDetails in the DatabaseConnector package.

xSpecCohort   The number of the "extremely specific (xSpec)" cohort definition id in the cohort table (for noisy positives)

cdmDatabaseSchema

> The name of the database schema that contains the OMOP CDM instance. Requires read permissions to this database. On SQL Server, this should specifiy 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.

cohortDatabaseTable

> 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.

outDatabaseSchema
              The name of a database schema where the user has write capability. A temporary
              cohort table will be created here.

modelOutputFileName
              A string designation for the training model file

xSensCohort   The number of the "extremely sensitive (xSens)" cohort definition id in the co-
                  hort table (used to exclude subjects from the base population)

prevalenceCohort
              The number of the cohort definition id to determine the disease prevalence, (de-
              fault=xSensCohort)

excludedConcepts
              A list of conceptIds to exclude from featureExtraction. These should include all
              concept_ids that were used to define the xSpec model (default=NULL)

addDescendantsToExclude
              Should descendants of excluded concepts also be excluded? (default=FALSE)

mainPopulationCohort
              The number of the cohort ID to be used as a base population for the model
              (default=NULL)

lowerAgeLimit   The lower age for subjects in the model (default=NULL)

upperAgeLimit   The upper age for subjects in the model (default=NULL)

gender   The gender(s) to be included (default c(8507, 8532))

startDate   The starting date for including subjects in the model (default=NULL)

endDate   The ending date for including subjects in the model (default=NULL)

cdmVersion   The CDM version of the database (default=5)

outFolder   The folder where the output files will be written (default=working directory)

Details

Function to create a diagnostic prediction model for a health outcome of interest using the xSpec
cohort. The model may be applied to the evaluation cohort to determine probabilities for each
subject for the health outcome of interest.

PheValuator   *PheValuator*

Description

PheValuator

---

testPhenotypeAlgorithm

Test phenotype algorithms

Description

Test phenotype algorithms

Usage

```
testPhenotypeAlgorithm(connectionDetails, cutPoints = c(0.1, 0.2, 0.3,
  0.4, 0.5, "EV", 0.6, 0.7, 0.8, 0.9), evaluationOutputFileName,
  phenotypeCohortId, cdmShortName = "", phenotypeText = "",
  order = 1, modelText = "", xSpecCohort = "", xSensCohort = "",
  prevalenceCohort = "", cohortDatabaseSchema, cohortTable)
```

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

evaluationOutputFileName
: The full file name with path for the evaluation file

phenotypeCohortId
: The number of the cohort of the phenotype algorithm to test

cdmShortName
: A string to identify the CDM tested (Default = NULL)

phenotypeText
: A string to identify the phenotype algorithm in the output file

order
: The order of this algorithm for sorting in the output file (used when there are multiple phenotypes to test) (Default = 1)

modelText
: Descriptive name for the model (Default = NULL)

xSpecCohort
: The number of the "extremely specific (xSpec)" cohort definition id in the cohort table (for noisy positives) (Default = NULL)

xSensCohort
: The number of the "extremely sensitive (xSens)" cohort definition id in the cohort table (used to exclude subjects from the base population) (Default = NULL)

prevalenceCohort
: The number of the cohort definition id to determine the disease prevalence, (default=xSensCohort)

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.

Details

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

Value

A list containing 2 dataframes: 1) results - a dataframe with the results from the phenotype algorithm evaluation 2) misses - a dataframe with a sample of subject ids for TPs, FPs, TNs, and FNs for the 50 percent and over prediction threshold

Index