Package 'FeatureExtraction'

October 23, 2017

Type Package

Title Generating Features for a Cohort

Version 2.0.0

Date 2017-06-01

Author Martijn J. Schuemie [aut, cre],
Marc A. Suchard [aut],
Patrick B. Ryan [aut],
Jenna Reps [aut]

Maintainer Martijn J. Schuemie <schuemie@ohdsi.org>

Description An R package for generating features (covariates) for a cohort using data in the Common Data Model.

License Apache License 2.0

Depends R (>= 3.2.2),
DatabaseConnector (>= 2.0.0),

Imports methods,
bit,
ff,
ffbase (>= 0.12.1),
plyr,
Rcpp (>= 0.11.2),
rJava,
jsonlite,
SqlRender (>= 1.1.3),

Suggests testthat,
knitr,
rmarkdown

LinkingTo Rcpp

NeedsCompilation yes

RoxygenNote 6.0.1

R topics documented:

`aggregateCovariates` *Aggregate covariate data*

Description

Aggregate covariate data

Usage

```
aggregateCovariates(covariateData)
```

Arguments

`covariateData`	An object of type `covariateData` as generated using `getDbCovariateData`.

Value

An object of class `covariateData`.

byMaxFf *Compute max of values binned by a second variable*

Description

Compute max of values binned by a second variable

Usage

```
byMaxFf(values, bins)
```

Arguments

values	An ff object containing the numeric values to take the max of.
bins	An ff object containing the numeric values to bin by.

Examples

```
values <- ff::as.ff(c(1, 1, 2, 2, 1))
bins <- ff::as.ff(c(1, 1, 1, 2, 2))
byMaxFf(values, bins)
```

bySumFf *Compute sum of values binned by a second variable*

Description

Compute sum of values binned by a second variable

Usage

```
bySumFf(values, bins)
```

Arguments

values	An ff object containing the numeric values to take the sum of.
bins	An ff object containing the numeric values to bin by.

Examples

```
values <- ff::as.ff(c(1, 1, 2, 2, 1))
bins <- ff::as.ff(c(1, 1, 1, 2, 2))
bySumFf(values, bins)
```

`computeStandardizedDifference` *Compute standardized difference of mean for all covariates.*

Description

Computes the standardized difference for all covariates between two cohorts. The standardized difference is defined as the difference between the mean divided by the overall standard deviation.

Usage

```
computeStandardizedDifference(covariateData1, covariateData2)
```

Arguments

`covariateData1`	The covariate data of the first cohort. Needs to be in aggregated format.
`covariateData2`	The covariate data of the second cohort. Needs to be in aggregated format.

Value

A data frame with means and standard deviations per cohort as well as the standardized difference of mean.

`convertPrespecSettingsToDetailedSettings` *Convert prespecified covariate settings into detailed covariate settings*

Description

Convert prespecified covariate settings into detailed covariate settings

Usage

```
convertPrespecSettingsToDetailedSettings(covariateSettings)
```

Arguments

`covariateSettings`	An object of type `covariateSettings` as created for example by the `createCovariateSettings` function.

Details

For advanced users only.

Value

An object of type `covariateSettings`, to be used in other functions.

createAnalysisDetails *Create detailed covariate settings*

Description

Create detailed covariate settings

Usage

```
createAnalysisDetails(analysisId, sqlFileName, parameters,
  includedCovariateConceptIds = c(), addDescendantsToInclude = FALSE,
  excludedCovariateConceptIds = c(), addDescendantsToExclude = FALSE,
  includedCovariateIds = c())
```

Arguments

analysisId	An integer between 0 and 999 that uniquely identifies this analysis.
sqlFileName	The name of the paramterized SQL file embedded in the featureExtraction package.
parameters	The list of parameter values used to render the template SQL.
includedCovariateConceptIds	A list of concept IDs that should be used to construct covariates.
addDescendantsToInclude	Should descendant concept IDs be added to the list of concepts to include?
excludedCovariateConceptIds	A list of concept IDs that should NOT be used to construct covariates.
addDescendantsToExclude	Should descendant concept IDs be added to the list of concepts to exclude?
includedCovariateIds	A list of covariate IDs that should be restricted to.

Details

creates an object specifying in detail how covariates should be contructed from data in the CDM model. Warning: this function is for advanced users only.

Value

An object of type analysisDetail, to be used in createDetailedCovariateSettings or createDetailedTemporalC

Examples

```
analysisDetails <- createAnalysisDetails(analysisId = 1,
                                         sqlFileName = "DemographicsGender.sql",
                                         parameters = list(analysisId = 1,
                                                           analysisName = "Gender",
                                                           domainId = "Demographics"),
                                         includedCovariateConceptIds = c(),
                                         addDescendantsToInclude = FALSE,
                                         excludedCovariateConceptIds = c(),
                                         addDescendantsToExclude = FALSE,
```

```
                                    includedCovariateIds = c())
```

`createCohortAttrCovariateSettings` *Create cohort attribute covariate settings*

Description

Create cohort attribute covariate settings

Usage

```
createCohortAttrCovariateSettings(attrDatabaseSchema,
  attrDefinitionTable = "attribute_definition",
  cohortAttrTable = "cohort_attribute", includeAttrIds = c())
```

Arguments

`attrDatabaseSchema`	The database schema where the attribute definition and cohort attribute table can be found.
`attrDefinitionTable`	The name of the attribute definition table.
`cohortAttrTable`	The name of the cohort attribute table.
`includeAttrIds`	(optional) A list of attribute definition IDs to restrict to.

Details

Creates an object specifying where the cohort attributes can be found to construct covariates. The attributes should be defined in a table with the same structure as the attribute_definition table in the Common Data Model. It should at least have these columns:

attribute_definition_id A unique identifier of type integer.

attribute_name A short description of the attribute.

The cohort attributes themselves should be stored in a table with the same format as the cohort_attribute table in the Common Data Model. It should at least have these columns:

cohort_definition_id A key to link to the cohort table.

subject_id A key to link to the cohort table.

cohort_start_date A key to link to the cohort table.

attribute_definition_id An foreign key linking to the attribute definition table.

value_as_number A real number.

Value

An object of type `covariateSettings`, to be used in other functions.

createCovariateSettings *Create covariate settings*

Description

Create covariate settings

Usage

```
createCovariateSettings(useDemographicsGender = FALSE,
  useDemographicsAge = FALSE, useDemographicsAgeGroup = FALSE,
  useDemographicsRace = FALSE, useDemographicsEthnicity = FALSE,
  useDemographicsIndexYear = FALSE, useDemographicsIndexMonth = FALSE,
  useDemographicsPriorObservationTime = FALSE,
  useDemographicsPostObservationTime = FALSE,
  useDemographicsTimeInCohort = FALSE,
  useConditionOccurrenceAnyTimePrior = FALSE,
  useConditionOccurrenceLongTerm = FALSE,
  useConditionOccurrenceMediumTerm = FALSE,
  useConditionOccurrenceShortTerm = FALSE,
  useConditionOccurrenceInpatientAnyTimePrior = FALSE,
  useConditionOccurrenceInpatientLongTerm = FALSE,
  useConditionOccurrenceInpatientMediumTerm = FALSE,
  useConditionOccurrenceInpatientShortTerm = FALSE,
  useConditionEraAnyTimePrior = FALSE, useConditionEraLongTerm = FALSE,
  useConditionEraMediumTerm = FALSE, useConditionEraShortTerm = FALSE,
  useConditionEraOverlapping = FALSE, useConditionEraStartLongTerm = FALSE,
  useConditionEraStartMediumTerm = FALSE,
  useConditionEraStartShortTerm = FALSE,
  useConditionGroupEraAnyTimePrior = FALSE,
  useConditionGroupEraLongTerm = FALSE,
  useConditionGroupEraMediumTerm = FALSE,
  useConditionGroupEraShortTerm = FALSE,
  useConditionGroupEraOverlapping = FALSE,
  useConditionGroupEraStartLongTerm = FALSE,
  useConditionGroupEraStartMediumTerm = FALSE,
  useConditionGroupEraStartShortTerm = FALSE,
  useDrugExposureAnyTimePrior = FALSE, useDrugExposureLongTerm = FALSE,
  useDrugExposureMediumTerm = FALSE, useDrugExposureShortTerm = FALSE,
  useDrugEraAnyTimePrior = FALSE, useDrugEraLongTerm = FALSE,
  useDrugEraMediumTerm = FALSE, useDrugEraShortTerm = FALSE,
  useDrugEraOverlapping = FALSE, useDrugEraStartLongTerm = FALSE,
  useDrugEraStartMediumTerm = FALSE, useDrugEraStartShortTerm = FALSE,
  useDrugGroupEraAnyTimePrior = FALSE, useDrugGroupEraLongTerm = FALSE,
  useDrugGroupEraMediumTerm = FALSE, useDrugGroupEraShortTerm = FALSE,
  useDrugGroupEraOverlapping = FALSE, useDrugGroupEraStartLongTerm = FALSE,
  useDrugGroupEraStartMediumTerm = FALSE,
  useDrugGroupEraStartShortTerm = FALSE,
  useProcedureOccurrenceAnyTimePrior = FALSE,
  useProcedureOccurrenceLongTerm = FALSE,
```

```
  useProcedureOccurrenceMediumTerm = FALSE,
  useProcedureOccurrenceShortTerm = FALSE,
  useDeviceExposureAnyTimePrior = FALSE, useDeviceExposureLongTerm = FALSE,
  useDeviceExposureMediumTerm = FALSE, useDeviceExposureShortTerm = FALSE,
  useMeasurementAnyTimePrior = FALSE, useMeasurementLongTerm = FALSE,
  useMeasurementMediumTerm = FALSE, useMeasurementShortTerm = FALSE,
  useMeasurementValueAnyTimePrior = FALSE,
  useMeasurementValueLongTerm = FALSE,
  useMeasurementValueMediumTerm = FALSE,
  useMeasurementValueShortTerm = FALSE,
  useMeasurementRangeGroupAnyTimePrior = FALSE,
  useMeasurementRangeGroupLongTerm = FALSE,
  useMeasurementRangeGroupMediumTerm = FALSE,
  useMeasurementRangeGroupShortTerm = FALSE,
  useObservationAnyTimePrior = FALSE, useObservationLongTerm = FALSE,
  useObservationMediumTerm = FALSE, useObservationShortTerm = FALSE,
  useCharlsonIndex = FALSE, useDcsi = FALSE, useChads2 = FALSE,
  useChads2Vasc = FALSE, useDistinctConditionCountLongTerm = FALSE,
  useDistinctConditionCountMediumTerm = FALSE,
  useDistinctConditionCountShortTerm = FALSE,
  useDistinctIngredientCountLongTerm = FALSE,
  useDistinctIngredientCountMediumTerm = FALSE,
  useDistinctIngredientCountShortTerm = FALSE,
  useDistinctProcedureCountLongTerm = FALSE,
  useDistinctProcedureCountMediumTerm = FALSE,
  useDistinctProcedureCountShortTerm = FALSE,
  useDistinctMeasurementCountLongTerm = FALSE,
  useDistinctMeasurementCountMediumTerm = FALSE,
  useDistinctMeasurementCountShortTerm = FALSE,
  useVisitCountLongTerm = FALSE, useVisitCountMediumTerm = FALSE,
  useVisitCountShortTerm = FALSE, longTermStartDays = -365,
  mediumTermStartDays = -180, shortTermStartDays = -30, endDays = 0,
  includedCovariateConceptIds = c(), addDescendantsToInclude = FALSE,
  excludedCovariateConceptIds = c(), addDescendantsToExclude = FALSE,
  includedCovariateIds = c())
```

Arguments

`useDemographicsGender`
: Gender of the subject. (analysis ID 1)

`useDemographicsAge`
: Age of the subject on the index date (in years). (analysis ID 2)

`useDemographicsAgeGroup`
: Age of the subject on the index date (in 5 year age groups) (analysis ID 3)

`useDemographicsRace`
: Race of the subject. (analysis ID 4)

`useDemographicsEthnicity`
: Ethnicity of the subject. (analysis ID 5)

`useDemographicsIndexYear`
: Year of the index date. (analysis ID 6)

`useDemographicsIndexMonth`
: Month of the index date. (analysis ID 7)

useDemographicsPriorObservationTime
: Number of days of observation time preceding the index date. (analysis ID 8)

useDemographicsPostObservationTime
: Number of days of observation time preceding the index date. (analysis ID 9)

useDemographicsTimeInCohort
: Number of days of observation time preceding the index date. (analysis ID 10)

useConditionOccurrenceAnyTimePrior
: One covariate per condition in the condition_occurrence table starting any time prior to index. (analysis ID 101)

useConditionOccurrenceLongTerm
: One covariate per condition in the condition_occurrence table starting in the long term window. (analysis ID 102)

useConditionOccurrenceMediumTerm
: One covariate per condition in the condition_occurrence table starting in the medium term window. (analysis ID 103)

useConditionOccurrenceShortTerm
: One covariate per condition in the condition_occurrence table starting in the short term window. (analysis ID 104)

useConditionOccurrenceInpatientAnyTimePrior
: One covariate per condition observed in an inpatient setting in the condition_occurrence table starting any time prior to index. (analysis ID 105)

useConditionOccurrenceInpatientLongTerm
: One covariate per condition observed in an inpatient setting in the condition_occurrence table starting in the long term window. (analysis ID 106)

useConditionOccurrenceInpatientMediumTerm
: One covariate per condition observed in an inpatient setting in the condition_occurrence table starting in the medium term window. (analysis ID 107)

useConditionOccurrenceInpatientShortTerm
: One covariate per condition observed in an inpatient setting in the condition_occurrence table starting in the short term window. (analysis ID 108)

useConditionEraAnyTimePrior
: One covariate per condition in the condition_era table overlapping with any time prior to index. (analysis ID 201)

useConditionEraLongTerm
: One covariate per condition in the condition_era table overlapping with any part of the long term window. (analysis ID 202)

useConditionEraMediumTerm
: One covariate per condition in the condition_era table overlapping with any part of the medium term window. (analysis ID 203)

useConditionEraShortTerm
: One covariate per condition in the condition_era table overlapping with any part of the short term window. (analysis ID 204)

useConditionEraOverlapping
: One covariate per condition in the condition_era table overlapping with the end of the risk window. (analysis ID 205)

useConditionEraStartLongTerm
: One covariate per condition in the condition_era table starting in the long term window. (analysis ID 206)

useConditionEraStartMediumTerm
: One covariate per condition in the condition_era table starting in the medium term window. (analysis ID 207)

useConditionEraStartShortTerm
: One covariate per condition in the condition_era table starting in the short term window. (analysis ID 208)

useConditionGroupEraAnyTimePrior
: One covariate per condition era rolled up to groups in the condition_era table overlapping with any time prior to index. (analysis ID 209)

useConditionGroupEraLongTerm
: One covariate per condition era rolled up to groups in the condition_era table overlapping with any part of the long term window. (analysis ID 210)

useConditionGroupEraMediumTerm
: One covariate per condition era rolled up to groups in the condition_era table overlapping with any part of the medium term window. (analysis ID 211)

useConditionGroupEraShortTerm
: One covariate per condition era rolled up to groups in the condition_era table overlapping with any part of the short term window. (analysis ID 212)

useConditionGroupEraOverlapping
: One covariate per condition era rolled up to groups in the condition_era table overlapping with the end of the risk window. (analysis ID 213)

useConditionGroupEraStartLongTerm
: One covariate per condition era rolled up to groups in the condition_era table starting in the long term window. (analysis ID 214)

useConditionGroupEraStartMediumTerm
: One covariate per condition era rolled up to groups in the condition_era table starting in the medium term window. (analysis ID 215)

useConditionGroupEraStartShortTerm
: One covariate per condition era rolled up to groups in the condition_era table starting in the short term window. (analysis ID 216)

useDrugExposureAnyTimePrior
: One covariate per drug in the drug_exposure table starting any time prior to index. (analysis ID 301)

useDrugExposureLongTerm
: One covariate per drug in the drug_exposure table starting in the long term window. (analysis ID 302)

useDrugExposureMediumTerm
: One covariate per drug in the drug_exposure table starting in the medium term window. (analysis ID 303)

useDrugExposureShortTerm
: One covariate per drug in the drug_exposure table starting in the short term window. (analysis ID 304)

useDrugEraAnyTimePrior
: One covariate per drug in the drug_era table overlapping with any time prior to index. (analysis ID 401)

useDrugEraLongTerm
: One covariate per drug in the drug_era table overlapping with any part of the long term window. (analysis ID 402)

useDrugEraMediumTerm
: One covariate per drug in the drug_era table overlapping with any part of the medium term window. (analysis ID 403)

useDrugEraShortTerm
: One covariate per drug in the drug_era table overlapping with any part of the short window. (analysis ID 404)

useDrugEraOverlapping
: One covariate per drug in the drug_era table overlapping with the end of the risk window. (analysis ID 405)

useDrugEraStartLongTerm
: One covariate per drug in the drug_era table starting in the long term window. (analysis ID 406)

useDrugEraStartMediumTerm
: One covariate per drug in the drug_era table starting in the medium term window. (analysis ID 407)

useDrugEraStartShortTerm
: One covariate per drug in the drug_era table starting in the long short window. (analysis ID 408)

useDrugGroupEraAnyTimePrior
: One covariate per drug rolled up to ATC groups in the drug_era table overlapping with any time prior to index. (analysis ID 409)

useDrugGroupEraLongTerm
: One covariate per drug rolled up to ATC groups in the drug_era table overlapping with any part of the long term window. (analysis ID 410)

useDrugGroupEraMediumTerm
: One covariate per drug rolled up to ATC groups in the drug_era table overlapping with any part of the medium term window. (analysis ID 411)

useDrugGroupEraShortTerm
: One covariate per drug rolled up to ATC groups in the drug_era table overlapping with any part of the short term window. (analysis ID 412)

useDrugGroupEraOverlapping
: One covariate per drug rolled up to ATC groups in the drug_era table overlapping with the end of the risk window. (analysis ID 413)

useDrugGroupEraStartLongTerm
: One covariate per drug rolled up to ATC groups in the drug_era table starting in the long term window. (analysis ID 414)

useDrugGroupEraStartMediumTerm
: One covariate per drug rolled up to ATC groups in the drug_era table starting in the medium term window. (analysis ID 415)

useDrugGroupEraStartShortTerm
: One covariate per drug rolled up to ATC groups in the drug_era table starting in the short term window. (analysis ID 416)

useProcedureOccurrenceAnyTimePrior
: One covariate per procedure in the procedure_occurrence table any time prior to index. (analysis ID 501)

useProcedureOccurrenceLongTerm
: One covariate per procedure in the procedure_occurrence table in the long term window. (analysis ID 502)

useProcedureOccurrenceMediumTerm
: One covariate per procedure in the procedure_occurrence table in the medium term window. (analysis ID 503)

useProcedureOccurrenceShortTerm
: One covariate per procedure in the procedure_occurrence table in the short term window. (analysis ID 504)

useDeviceExposureAnyTimePrior
: One covariate per device in the device exposure table starting any time prior to index. (analysis ID 601)

useDeviceExposureLongTerm
: One covariate per device in the device exposure table starting in the long term window. (analysis ID 602)

useDeviceExposureMediumTerm
: One covariate per device in the device exposure table starting in the medium term window. (analysis ID 603)

useDeviceExposureShortTerm
: One covariate per device in the device exposure table starting in the short term window. (analysis ID 604)

useMeasurementAnyTimePrior
: One covariate per measurement in the measurement table any time prior to index. (analysis ID 701)

useMeasurementLongTerm
: One covariate per measurement in the measurement table in the long term window. (analysis ID 702)

useMeasurementMediumTerm
: One covariate per measurement in the measurement table in the medium term window. (analysis ID 703)

useMeasurementShortTerm
: One covariate per measurement in the measurement table in the short term window. (analysis ID 704)

useMeasurementValueAnyTimePrior
: One covariate containing the value per measurement-unit combination any time prior to index. (analysis ID 705)

useMeasurementValueLongTerm
: One covariate containing the value per measurement-unit combination in the long term window. (analysis ID 706)

useMeasurementValueMediumTerm
: One covariate containing the value per measurement-unit combination in the medium term window. (analysis ID 707)

useMeasurementValueShortTerm
: One covariate containing the value per measurement-unit combination in the short term window. (analysis ID 708)

useMeasurementRangeGroupAnyTimePrior
: Covariates indicating whether measurements are below, within, or above normal range any time prior to index. (analysis ID 709)

useMeasurementRangeGroupLongTerm
: Covariates indicating whether measurements are below, within, or above normal range in the long term window. (analysis ID 710)

useMeasurementRangeGroupMediumTerm
: Covariates indicating whether measurements are below, within, or above normal range in the medium term window. (analysis ID 711)

useMeasurementRangeGroupShortTerm
: Covariates indicating whether measurements are below, within, or above normal range in the short term window. (analysis ID 712)

useObservationAnyTimePrior
: One covariate per observation in the observation table any time prior to index. (analysis ID 801)

useObservationLongTerm
: One covariate per observation in the observation table in the long term window. (analysis ID 802)

useObservationMediumTerm
: One covariate per observation in the observation table in the medium term window. (analysis ID 803)

useObservationShortTerm
: One covariate per observation in the observation table in the short term window. (analysis ID 804)

useCharlsonIndex
: The Charlson comorbidity index (Romano adaptation) using all conditions prior to the window end. (analysis ID 901)

useDcsi
: The Diabetes Comorbidity Severity Index (DCSI) using all conditions prior to the window end. (analysis ID 902)

useChads2
: The CHADS2 score using all conditions prior to the window end. (analysis ID 903)

useChads2Vasc
: The CHADS2VASc score using all conditions prior to the window end. (analysis ID 904)

useDistinctConditionCountLongTerm
: The number of distinct condition concepts observed in the long term window. (analysis ID 905)

useDistinctConditionCountMediumTerm
: The number of distinct condition concepts observed in the medium term window. (analysis ID 906)

useDistinctConditionCountShortTerm
: The number of distinct condition concepts observed in the medium term window. (analysis ID 907)

useDistinctIngredientCountLongTerm
: The number of distinct ingredients observed in the long term window. (analysis ID 908)

useDistinctIngredientCountMediumTerm
: The number of distinct ingredients observed in the medium term window. (analysis ID 909)

useDistinctIngredientCountShortTerm
: The number of distinct ingredients observed in the medium term window. (analysis ID 910)

useDistinctProcedureCountLongTerm
: The number of distinct procedures observed in the long term window. (analysis ID 911)

useDistinctProcedureCountMediumTerm
: The number of distinct procedures observed in the medium term window. (analysis ID 912)

useDistinctProcedureCountShortTerm
: The number of distinct procedures observed in the medium term window. (analysis ID 913)

useDistinctMeasurementCountLongTerm
: The number of distinct measurements observed in the long term window. (analysis ID 914)

useDistinctMeasurementCountMediumTerm
: The number of distinct measurements observed in the medium term window. (analysis ID 915)

useDistinctMeasurementCountShortTerm
: The number of distinct measurements observed in the medium term window. (analysis ID 916)

useVisitCountLongTerm
: The number of visits observed in the long term window. (analysis ID 917)

useVisitCountMediumTerm
: The number of visits observed in the medium term window. (analysis ID 918)

useVisitCountShortTerm
: The number of visits observed in the medium term window. (analysis ID 919)

longTermStartDays
: What is the start day (relative to the index date) of the long-term window?

mediumTermStartDays
: What is the start day (relative to the index date) of the medium-term window?

shortTermStartDays
: What is the start day (relative to the index date) of the short-term window?

endDays
: What is the end day (relative to the index date) of the window?

includedCovariateConceptIds
: A list of concept IDs that should be used to construct covariates.

addDescendantsToInclude
: Should descendant concept IDs be added to the list of concepts to include?

excludedCovariateConceptIds
: A list of concept IDs that should NOT be used to construct covariates.

addDescendantsToExclude
: Should descendant concept IDs be added to the list of concepts to exclude?

includedCovariateIds
: A list of covariate IDs that should be restricted to.

Details

creates an object specifying how covariates should be contructed from data in the CDM model.

Value

An object of type `covariateSettings`, to be used in other functions.

Examples

```
settings <- createCovariateSettings(useDemographicsGender = TRUE,
                                    useDemographicsAge = FALSE,
                                    useDemographicsAgeGroup = TRUE,
                                    useDemographicsRace = TRUE,
                                    useDemographicsEthnicity = TRUE,
                                    useDemographicsIndexYear = TRUE,
                                    useDemographicsIndexMonth = TRUE,
                                    useDemographicsPriorObservationTime = FALSE,
                                    useDemographicsPostObservationTime = FALSE,
                                    useDemographicsTimeInCohort = FALSE,
                                    useConditionOccurrenceAnyTimePrior = FALSE,
                                    useConditionOccurrenceLongTerm = FALSE,
```

```
                                    useConditionOccurrenceMediumTerm = FALSE,
                                    useConditionOccurrenceShortTerm = FALSE,
                                    useConditionOccurrenceInpatientAnyTimePrior = FALSE,
                                    useConditionOccurrenceInpatientLongTerm = FALSE,
                                    useConditionOccurrenceInpatientMediumTerm = FALSE,
                                    useConditionOccurrenceInpatientShortTerm = FALSE,
                                    useConditionEraAnyTimePrior = FALSE,
                                    useConditionEraLongTerm = FALSE,
                                    useConditionEraMediumTerm = FALSE,
                                    useConditionEraShortTerm = FALSE,
                                    useConditionEraOverlapping = FALSE,
                                    useConditionEraStartLongTerm = FALSE,
                                    useConditionEraStartMediumTerm = FALSE,
                                    useConditionEraStartShortTerm = FALSE,
                                    useConditionGroupEraAnyTimePrior = FALSE,
                                    useConditionGroupEraLongTerm = TRUE,
                                    useConditionGroupEraMediumTerm = FALSE,
                                    useConditionGroupEraShortTerm = TRUE,
                                    useConditionGroupEraOverlapping = FALSE,
                                    useConditionGroupEraStartLongTerm = FALSE,
                                    useConditionGroupEraStartMediumTerm = FALSE,
                                    useConditionGroupEraStartShortTerm = FALSE,
                                    useDrugExposureAnyTimePrior = FALSE,
                                    useDrugExposureLongTerm = FALSE,
                                    useDrugExposureMediumTerm = FALSE,
                                    useDrugExposureShortTerm = FALSE,
                                    useDrugEraAnyTimePrior = FALSE,
                                    useDrugEraLongTerm = FALSE,
                                    useDrugEraMediumTerm = FALSE,
                                    useDrugEraShortTerm = FALSE,
                                    useDrugEraOverlapping = FALSE,
                                    useDrugEraStartLongTerm = FALSE,
                                    useDrugEraStartMediumTerm = FALSE,
                                    useDrugEraStartShortTerm = FALSE,
                                    useDrugGroupEraAnyTimePrior = FALSE,
                                    useDrugGroupEraLongTerm = TRUE,
                                    useDrugGroupEraMediumTerm = FALSE,
                                    useDrugGroupEraShortTerm = TRUE,
                                    useDrugGroupEraOverlapping = TRUE,
                                    useDrugGroupEraStartLongTerm = FALSE,
                                    useDrugGroupEraStartMediumTerm = FALSE,
                                    useDrugGroupEraStartShortTerm = FALSE,
                                    useProcedureOccurrenceAnyTimePrior = FALSE,
                                    useProcedureOccurrenceLongTerm = TRUE,
                                    useProcedureOccurrenceMediumTerm = FALSE,
                                    useProcedureOccurrenceShortTerm = TRUE,
                                    useDeviceExposureAnyTimePrior = FALSE,
                                    useDeviceExposureLongTerm = TRUE,
                                    useDeviceExposureMediumTerm = FALSE,
                                    useDeviceExposureShortTerm = TRUE,
                                    useMeasurementAnyTimePrior = FALSE,
                                    useMeasurementLongTerm = TRUE,
                                    useMeasurementMediumTerm = FALSE,
                                    useMeasurementShortTerm = TRUE,
                                    useMeasurementValueAnyTimePrior = FALSE,
                                    useMeasurementValueLongTerm = FALSE,
                                    useMeasurementValueMediumTerm = FALSE,
```

```
                                    useMeasurementValueShortTerm = FALSE,
                                    useMeasurementRangeGroupAnyTimePrior = FALSE,
                                    useMeasurementRangeGroupLongTerm = TRUE,
                                    useMeasurementRangeGroupMediumTerm = FALSE,
                                    useMeasurementRangeGroupShortTerm = FALSE,
                                    useObservationAnyTimePrior = FALSE,
                                    useObservationLongTerm = TRUE,
                                    useObservationMediumTerm = FALSE,
                                    useObservationShortTerm = TRUE,
                                    useCharlsonIndex = TRUE,
                                    useDcsi = TRUE,
                                    useChads2 = TRUE,
                                    useChads2Vasc = TRUE,
                                    useDistinctConditionCountLongTerm = FALSE,
                                    useDistinctConditionCountMediumTerm = FALSE,
                                    useDistinctConditionCountShortTerm = FALSE,
                                    useDistinctIngredientCountLongTerm = FALSE,
                                    useDistinctIngredientCountMediumTerm = FALSE,
                                    useDistinctIngredientCountShortTerm = FALSE,
                                    useDistinctProcedureCountLongTerm = FALSE,
                                    useDistinctProcedureCountMediumTerm = FALSE,
                                    useDistinctProcedureCountShortTerm = FALSE,
                                    useDistinctMeasurementCountLongTerm = FALSE,
                                    useDistinctMeasurementCountMediumTerm = FALSE,
                                    useDistinctMeasurementCountShortTerm = FALSE,
                                    useVisitCountLongTerm = FALSE,
                                    useVisitCountMediumTerm = FALSE,
                                    useVisitCountShortTerm = FALSE,
                                    longTermStartDays = -365,
                                    mediumTermStartDays = -180,
                                    shortTermStartDays = -30,
                                    endDays = 0,
                                    includedCovariateConceptIds = c(),
                                    addDescendantsToInclude = FALSE,
                                    excludedCovariateConceptIds = c(),
                                    addDescendantsToExclude = FALSE,
                                    includedCovariateIds = c())
```

`createDefaultCovariateSettings` *Create default covariate settings*

Description

Create default covariate settings

Usage

```
createDefaultCovariateSettings(includedCovariateConceptIds = c(),
  addDescendantsToInclude = FALSE, excludedCovariateConceptIds = c(),
  addDescendantsToExclude = FALSE, includedCovariateIds = c())
```

Arguments

includedCovariateConceptIds
: A list of concept IDs that should be used to construct covariates.

addDescendantsToInclude
: Should descendant concept IDs be added to the list of concepts to include?

excludedCovariateConceptIds
: A list of concept IDs that should NOT be used to construct covariates.

addDescendantsToExclude
: Should descendant concept IDs be added to the list of concepts to exclude?

includedCovariateIds
: A list of covariate IDs that should be restricted to.

Value

An object of type `covariateSettings`, to be used in other functions.

---

createDefaultTemporalCovariateSettings
Create default covariate settings

Description

Create default covariate settings

Usage

```
createDefaultTemporalCovariateSettings(includedCovariateConceptIds = c(),
  addDescendantsToInclude = FALSE, excludedCovariateConceptIds = c(),
  addDescendantsToExclude = FALSE, includedCovariateIds = c())
```

Arguments

includedCovariateConceptIds
: A list of concept IDs that should be used to construct covariates.

addDescendantsToInclude
: Should descendant concept IDs be added to the list of concepts to include?

excludedCovariateConceptIds
: A list of concept IDs that should NOT be used to construct covariates.

addDescendantsToExclude
: Should descendant concept IDs be added to the list of concepts to exclude?

includedCovariateIds
: A list of covariate IDs that should be restricted to.

Value

An object of type `covariateSettings`, to be used in other functions.

createDetailedCovariateSettings

Create detailed covariate settings

Description

Create detailed covariate settings

Usage

```
createDetailedCovariateSettings(analyses = list())
```

Arguments

analyses	A list of analysisDetail objects as created using createAnalysisDetails.

Details

creates an object specifying in detail how covariates should be contructed from data in the CDM model. Warning: this function is for advanced users only.

Value

An object of type covariateSettings, to be used in other functions.

createDetailedTemporalCovariateSettings

Create detailed temporal covariate settings

Description

Create detailed temporal covariate settings

Usage

```
createDetailedTemporalCovariateSettings(analyses = list(),
  temporalStartDays = -365:-1, temporalEndDays = -365:-1)
```

Arguments

analyses	A list of analysis detail objects as created using createAnalysisDetails.
temporalStartDays	A list of integers representing the start of a time period, relative to the index date. 0 indicates the index date, -1 indicates the day before the index date, etc. The start day is included in the time period.
temporalEndDays	A list of integers representing the end of a time period, relative to the index date. 0 indicates the index date, -1 indicates the day before the index date, etc. The end day is included in the time period.

Details

creates an object specifying in detail how temporal covariates should be contructed from data in the CDM model. Warning: this function is for advanced users only.

Value

An object of type `covariateSettings`, to be used in other functions.

`createHdpsCovariateSettings` *Create HDPS covariate settings*

Description

Create HDPS covariate settings

Usage

```
createHdpsCovariateSettings(useCovariateCohortIdIs1 = FALSE,
  useCovariateDemographics = TRUE, useCovariateDemographicsGender = TRUE,
  useCovariateDemographicsRace = TRUE,
  useCovariateDemographicsEthnicity = TRUE,
  useCovariateDemographicsAge = TRUE, useCovariateDemographicsYear = TRUE,
  useCovariateDemographicsMonth = TRUE,
  useCovariateConditionOccurrence = TRUE,
  useCovariate3DigitIcd9Inpatient180d = FALSE,
  useCovariate3DigitIcd9Inpatient180dMedF = FALSE,
  useCovariate3DigitIcd9Inpatient180d75F = FALSE,
  useCovariate3DigitIcd9Ambulatory180d = FALSE,
  useCovariate3DigitIcd9Ambulatory180dMedF = FALSE,
  useCovariate3DigitIcd9Ambulatory180d75F = FALSE,
  useCovariateDrugExposure = FALSE,
  useCovariateIngredientExposure180d = FALSE,
  useCovariateIngredientExposure180dMedF = FALSE,
  useCovariateIngredientExposure180d75F = FALSE,
  useCovariateProcedureOccurrence = FALSE,
  useCovariateProcedureOccurrenceInpatient180d = FALSE,
  useCovariateProcedureOccurrenceInpatient180dMedF = FALSE,
  useCovariateProcedureOccurrenceInpatient180d75F = FALSE,
  useCovariateProcedureOccurrenceAmbulatory180d = FALSE,
  useCovariateProcedureOccurrenceAmbulatory180dMedF = FALSE,
  useCovariateProcedureOccurrenceAmbulatory180d75F = FALSE,
  excludedCovariateConceptIds = c(), includedCovariateConceptIds = c(),
  deleteCovariatesSmallCount = 100)
```

Arguments

`useCovariateCohortIdIs1`
: A boolean value (TRUE/FALSE) to determine if a covariate should be contructed for whether the cohort ID is 1 (currently primarily used in CohortMethod).

useCovariateDemographics
: A boolean value (TRUE/FALSE) to determine if demographic covariates (age in 5-yr increments, gender, race, ethnicity, year of index date, month of index date) will be created and included in future models.

useCovariateDemographicsGender
: A boolean value (TRUE/FALSE) to determine if gender should be included in the model.

useCovariateDemographicsRace
: A boolean value (TRUE/FALSE) to determine if race should be included in the model.

useCovariateDemographicsEthnicity
: A boolean value (TRUE/FALSE) to determine if ethnicity should be included in the model.

useCovariateDemographicsAge
: A boolean value (TRUE/FALSE) to determine if age (in 5 year increments) should be included in the model.

useCovariateDemographicsYear
: A boolean value (TRUE/FALSE) to determine if calendar year should be included in the model.

useCovariateDemographicsMonth
: A boolean value (TRUE/FALSE) to determine if calendar month should be included in the model.

useCovariateConditionOccurrence
: A boolean value (TRUE/FALSE) to determine if covariates derived from CONDITION_OCCURRENCE table will be created and included in future models.

useCovariate3DigitIcd9Inpatient180d
: A boolean value (TRUE/FALSE) to determine if covariates will be created and used in models that look for presence/absence of condition within inpatient setting in 180d window prior to or on cohort index date. Conditions are aggregated at the ICD-9 3-digit level. Only applicable if useCovariateConditionOccurrence = TRUE.

useCovariate3DigitIcd9Inpatient180dMedF
: Similar to useCovariate3DigitIcd9Inpatient180d, but now only if the frequency of the ICD-9 code is higher than the median.

useCovariate3DigitIcd9Inpatient180d75F
: Similar to useCovariate3DigitIcd9Inpatient180d, but now only if the frequency of the ICD-9 code is higher than the 75th percentile.

useCovariate3DigitIcd9Ambulatory180d
: A boolean value (TRUE/FALSE) to determine if covariates will be created and used in models that look for presence/absence of condition within ambulatory setting in 180d window prior to or on cohort index date. Conditions are aggregated at the ICD-9 3-digit level. Only applicable if useCovariateConditionOccurrence = TRUE.

useCovariate3DigitIcd9Ambulatory180dMedF
: Similar to useCovariate3DigitIcd9Ambulatory180d, but now only if the frequency of the ICD-9 code is higher than the median.

useCovariate3DigitIcd9Ambulatory180d75F
: Similar to useCovariate3DigitIcd9Ambulatory180d, but now only if the frequency of the ICD-9 code is higher than the 75th percentile.

useCovariateDrugExposure
: A boolean value (TRUE/FALSE) to determine if covariates derived from DRUG_EXPOSURE table will be created and included in future models.

useCovariateIngredientExposure180d
: A boolean value (TRUE/FALSE) to determine if covariates will be created and used in models that look for presence/absence of drug ingredients within inpatient setting in 180d window prior to or on cohort index date. Only applicable if useCovariateDrugExposure = TRUE.

useCovariateIngredientExposure180dMedF
: Similar to useCovariateIngredientExposure180d, but now only if the frequency of the ingredient is higher than the median.

useCovariateIngredientExposure180d75F
: Similar to useCovariateIngredientExposure180d, but now only if the frequency of the ingredient is higher than the 75th percentile.

useCovariateProcedureOccurrence
: A boolean value (TRUE/FALSE) to determine if covariates derived from PROCEDURE_OCCURRENCE table will be created and included in future models.

useCovariateProcedureOccurrenceInpatient180d
: A boolean value (TRUE/FALSE) to determine if covariates will be created and used in models that look for presence/absence of procedures within inpatient setting in 180d window prior to or on cohort index date. Only applicable if useCovariateProcedureOccurrence = TRUE.

useCovariateProcedureOccurrenceInpatient180dMedF
: Similar to useCovariateProcedureOccurrenceInpatient180d, but now only if the frequency of the procedure code is higher than the median.

useCovariateProcedureOccurrenceInpatient180d75F
: Similar to useCovariateProcedureOccurrenceInpatient180d, but now only if the frequency of the procedure code is higher than the 75th percentile.

useCovariateProcedureOccurrenceAmbulatory180d
: A boolean value (TRUE/FALSE) to determine if covariates will be created and used in models that look for presence/absence of procedures within ambulatory setting in 180d window prior to or on cohort index date. Only applicable if useCovariateProcedureOccurrence = TRUE.

useCovariateProcedureOccurrenceAmbulatory180dMedF
: Similar to useCovariateProcedureOccurrenceAmbulatory180d, but now only if the frequency of the procedure code is higher than the median.

useCovariateProcedureOccurrenceAmbulatory180d75F
: Similar to useCovariateProcedureOccurrenceAmbulatory180d, but now only if the frequency of the procedure code is higher than the 75th percentile.

excludedCovariateConceptIds
: A list of concept IDs that should NOT be used to construct covariates.

includedCovariateConceptIds
: A list of concept IDs that should be used to construct covariates.

deleteCovariatesSmallCount
: A numeric value used to remove covariates that occur in both cohorts fewer than deleteCovariateSmallCounts time.

Details

creates an object specifying how covariates should be contructed from data in the CDM model.

Value

An object of type hdpsCovariateSettings, to be used in other functions.

createTable1 *Create a table 1*

Description

Creates a formatted table of cohort characteristics, to be included in publications or reports. Allows for creating a table describing a single cohort, or a table comparing two cohorts.

Usage

```
createTable1(covariateData1, covariateData2 = NULL,
  specifications = getDefaultTable1Specifications(), output = "two columns")
```

Arguments

covariateData1	The covariate data of the cohort to be included in the table.
covariateData2	The covariate data of the cohort to also be included, when comparing two cohorts.
specifications	Specifications of which covariates to display, and how.
output	The output format for the table. Options are `output = "two columns"`, `output = "one column"`, or `output = "list"`.

Value

A data frame, or, when `output = "list"` a list of two data frames.

createTable1CovariateSettings *Create covariate settings for a table 1*

Description

Creates a covariate settings object for generating only those covariates that will be included in a table 1. This function works by filtering the `covariateSettings` object for the covariates in the `specifications` object.

Usage

```
createTable1CovariateSettings(specifications = getDefaultTable1Specifications(),
 covariateSettings = createDefaultCovariateSettings(),
 includedCovariateConceptIds = c(), addDescendantsToInclude = FALSE,
 excludedCovariateConceptIds = c(), addDescendantsToExclude = FALSE,
 includedCovariateIds = c())
```

Arguments

specifications	A specifications object for generating a table using the createTable1 function.
covariateSettings	The covariate settings object to use as the basis for the filtered covariate settings.
includedCovariateConceptIds	A list of concept IDs that should be used to construct covariates.
addDescendantsToInclude	Should descendant concept IDs be added to the list of concepts to include?
excludedCovariateConceptIds	A list of concept IDs that should NOT be used to construct covariates.
addDescendantsToExclude	Should descendant concept IDs be added to the list of concepts to exclude?
includedCovariateIds	A list of covariate IDs that should be restricted to.

Value

A covariate settings object, for example to be used when calling the getDbCovariateData function.

---

createTemporalCovariateSettings
Create covariate settings

Description

Create covariate settings

Usage

```
createTemporalCovariateSettings(useDemographicsGender = FALSE,
  useDemographicsAge = FALSE, useDemographicsAgeGroup = FALSE,
  useDemographicsRace = FALSE, useDemographicsEthnicity = FALSE,
  useDemographicsIndexYear = FALSE, useDemographicsIndexMonth = FALSE,
  useDemographicsPriorObservationTime = FALSE,
  useDemographicsPostObservationTime = FALSE,
  useDemographicsTimeInCohort = FALSE, useConditionOccurrence = FALSE,
  useConditionOccurrenceInpatient = FALSE, useConditionEraStart = FALSE,
  useConditionEraOverlap = FALSE, useConditionEraGroupStart = FALSE,
  useConditionEraGroupOverlap = FALSE, useDrugExposure = FALSE,
  useDrugEraStart = FALSE, useDrugEraOverlap = FALSE,
  useDrugEraGroupStart = FALSE, useDrugEraGroupOverlap = FALSE,
  useProcedureOccurrence = FALSE, useDeviceExposure = FALSE,
  useMeasurement = FALSE, useMeasurementValue = FALSE,
  useMeasurementRangeGroup = FALSE, useObservation = FALSE,
  useCharlsonIndex = FALSE, useDcsi = FALSE, useChads2 = FALSE,
  useChads2Vasc = FALSE, useDistinctConditionCount = FALSE,
  useDistinctIngredientCount = FALSE, useDistinctProcedureCount = FALSE,
  useDistinctMeasurementCount = FALSE, useVisitCount = FALSE,
  temporalStartDays = -365:-1, temporalEndDays = -365:-1,
```

```
  includedCovariateConceptIds = c(), addDescendantsToInclude = FALSE,
  excludedCovariateConceptIds = c(), addDescendantsToExclude = FALSE,
  includedCovariateIds = c())
```

Arguments

useDemographicsGender
: Gender of the subject. (analysis ID 1)

useDemographicsAge
: Age of the subject on the index date (in years). (analysis ID 2)

useDemographicsAgeGroup
: Age of the subject on the index date (in 5 year age groups) (analysis ID 3)

useDemographicsRace
: Race of the subject. (analysis ID 4)

useDemographicsEthnicity
: Ethnicity of the subject. (analysis ID 5)

useDemographicsIndexYear
: Year of the index date. (analysis ID 6)

useDemographicsIndexMonth
: Month of the index date. (analysis ID 7)

useDemographicsPriorObservationTime
: Number of days of observation time preceding the index date. (analysis ID 8)

useDemographicsPostObservationTime
: Number of days of observation time preceding the index date. (analysis ID 9)

useDemographicsTimeInCohort
: Number of days of observation time preceding the index date. (analysis ID 10)

useConditionOccurrence
: One covariate per condition in the condition_occurrence table starting in the time window. (analysis ID 101)

useConditionOccurrenceInpatient
: One covariate per condition observed in an inpatient setting in the condition_occurrence table starting in the time window. (analysis ID 102)

useConditionEraStart
: One covariate per condition in the condition_era table starting in the time window. (analysis ID 201)

useConditionEraOverlap
: One covariate per condition in the condition_era table overlapping with any part of the time window. (analysis ID 202)

useConditionEraGroupStart
: One covariate per condition era rolled up to SNOMED groups in the condition_era table starting in the time window. (analysis ID 203)

useConditionEraGroupOverlap
: One covariate per condition era rolled up to SNOMED groups in the condition_era table overlapping with any part of the time window. (analysis ID 203)

useDrugExposure
: One covariate per drug in the drug_exposure table starting in the time window. (analysis ID 301)

useDrugEraStart
: One covariate per drug in the drug_era table starting in the time window. (analysis ID 401)

useDrugEraOverlap
: One covariate per drug in the drug_era table overlapping with any part of the time window. (analysis ID 402)

useDrugEraGroupStart
: One covariate per drug rolled up to ATC groups in the drug_era table starting in thetime window. (analysis ID 403)

useDrugEraGroupOverlap
: One covariate per drug rolled up to ATC groups in the drug_era table overlapping with any part of thetime window. (analysis ID 403)

useProcedureOccurrence
: One covariate per procedure in the procedure_occurrence table in the time window. (analysis ID 501)

useDeviceExposure
: One covariate per device in the device exposure table starting in the timewindow. (analysis ID 601)

useMeasurement
: One covariate per measurement in the measurement table in the time window. (analysis ID 701)

useMeasurementValue
: One covariate containing the value per measurement-unit combination in the time window. If multiple values are found, the last is taken. (analysis ID 702)

useMeasurementRangeGroup
: Covariates indicating whether measurements are below, within, or above normal range within the time period. (analysis ID 703)

useObservation
: One covariate per observation in the observation table in the time window. (analysis ID 801)

useCharlsonIndex
: The Charlson comorbidity index (Romano adaptation) using all conditions prior to the window end. (analysis ID 901)

useDcsi
: The Diabetes Comorbidity Severity Index (DCSI) using all conditions prior to the window end. (analysis ID 902)

useChads2
: The CHADS2 score using all conditions prior to the window end. (analysis ID 903)

useChads2Vasc
: The CHADS2VASc score using all conditions prior to the window end. (analysis ID 904)

useDistinctConditionCount
: The number of distinct condition concepts observed in the time window. (analysis ID 905)

useDistinctIngredientCount
: The number of distinct ingredients observed in the time window. (analysis ID 906)

useDistinctProcedureCount
: The number of distinct procedures observed in the time window. (analysis ID 907)

useDistinctMeasurementCount
: The number of distinct measurements observed in the time window. (analysis ID 908)

useVisitCount
: The number of visits observed in the time window. (analysis ID 909)

`temporalStartDays`
: A list of integers representing the start of a time period, relative to the index date. 0 indicates the index date, -1 indicates the day before the index date, etc. The start day is included in the time period.

`temporalEndDays`
: A list of integers representing the end of a time period, relative to the index date. 0 indicates the index date, -1 indicates the day before the index date, etc. The end day is included in the time period.

`includedCovariateConceptIds`
: A list of concept IDs that should be used to construct covariates.

`addDescendantsToInclude`
: Should descendant concept IDs be added to the list of concepts to include?

`excludedCovariateConceptIds`
: A list of concept IDs that should NOT be used to construct covariates.

`addDescendantsToExclude`
: Should descendant concept IDs be added to the list of concepts to exclude?

`includedCovariateIds`
: A list of covariate IDs that should be restricted to.

Details

creates an object specifying how covariates should be contructed from data in the CDM model.

Value

An object of type `covariateSettings`, to be used in other functions.

Examples

```
settings <- createTemporalCovariateSettings(useDemographicsGender = TRUE,
                                             useDemographicsAge = FALSE,
                                             useDemographicsAgeGroup = TRUE,
                                             useDemographicsRace = TRUE,
                                             useDemographicsEthnicity = TRUE,
                                             useDemographicsIndexYear = TRUE,
                                             useDemographicsIndexMonth = TRUE,
                                            useDemographicsPriorObservationTime = FALSE,
                                             useDemographicsPostObservationTime = FALSE,
                                             useDemographicsTimeInCohort = FALSE,
                                             useConditionOccurrence = FALSE,
                                             useConditionOccurrenceInpatient = FALSE,
                                             useConditionEraStart = FALSE,
                                             useConditionEraOverlap = FALSE,
                                             useConditionEraGroupStart = FALSE,
                                             useConditionEraGroupOverlap = TRUE,
                                             useDrugExposure = FALSE,
                                             useDrugEraStart = FALSE,
                                             useDrugEraOverlap = FALSE,
                                             useDrugEraGroupStart = FALSE,
                                             useDrugEraGroupOverlap = TRUE,
                                             useProcedureOccurrence = TRUE,
                                             useDeviceExposure = TRUE,
                                             useMeasurement = TRUE,
                                             useMeasurementValue = FALSE,
```

```
useMeasurementRangeGroup = TRUE,
useObservation = TRUE,
useCharlsonIndex = TRUE,
useDcsi = TRUE,
useChads2 = TRUE,
useChads2Vasc = TRUE,
useDistinctConditionCount = FALSE,
useDistinctIngredientCount = FALSE,
useDistinctProcedureCount = FALSE,
useDistinctMeasurementCount = FALSE,
useVisitCount = FALSE,
temporalStartDays = -365:-1,
temporalEndDays = -365:-1,
includedCovariateConceptIds = c(),
addDescendantsToInclude = FALSE,
excludedCovariateConceptIds = c(),
addDescendantsToExclude = FALSE,
includedCovariateIds = c())
```

`FeatureExtraction` *FeatureExtraction*

Description

FeatureExtraction

`filterByRowId` *Filter covariates by row ID*

Description

Filter covariates by row ID

Usage

```
filterByRowId(object, rowIds)
```

Arguments

`object`	Either an object of type `covariateData`, or an ffdf object containing covariate values.
`rowIds`	A vector (or ff object) containing the rowIds to keep.

Value

Either an object of type `covariateData`, or an ffdf object containing covariate values. (depending on the type of the `object` argument.

`getDbCohortAttrCovariatesData` *Getcovariate information from the database through the cohort_attribute table*

Description

Constructs a large default set of covariates for one or more cohorts using data in the CDM schema. Includes covariates for all drugs, drug classes, condition, condition classes, procedures, observations, etc.

Usage

```
getDbCohortAttrCovariatesData(connection, oracleTempSchema = NULL,
  cdmDatabaseSchema, cohortTable = "#cohort_person", cohortId = -1,
  rowIdField = "subject_id", covariateSettings, aggregated = FALSE)
```

Arguments

`connection`	A connection to the server containing the schema as created using the `connect` function in the `DatabaseConnector` package.
`oracleTempSchema`	A schema where temp tables can be created in Oracle.
`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'.
`cohortTable`	Name of the table holding the cohort for which we want to construct covariates. If it is a temp table, the name should have a hash prefix, e.g. '#temp_table'. If it is a non-temp table, it should include the database schema, e.g. 'cdm_database.cohort'.
`cohortId`	For which cohort ID should covariates be constructed? If set to -1, covariates will be constructed for all cohorts in the specified cohort table.
`rowIdField`	The name of the field in the cohort temp table that is to be used as the row_id field in the output table. This can be especially usefull if there is more than one period per person.
`covariateSettings`	An object of type `covariateSettings` as created using the `createCohortAttrCovariateSettings` function.
`aggregated`	Should aggregate statistics be computed instead of covariates per cohort entry?

Details

This function uses the data in the CDM to construct a large set of covariates for the provided cohort. The cohort is assumed to be in an existing temp table with these fields: 'subject_id', 'cohort_definition_id', 'cohort_start_date'. Optionally, an extra field can be added containing the unique identifier that will be used as rowID in the output. Typically, users don't call this function directly but rather use the `getDbCovariateData` function instead.

Value

Returns an object of type covariateData, containing information on the baseline covariates. Information about multiple outcomes can be captured at once for efficiency reasons. This object is a list with the following components:

covariates An ffdf object listing the baseline covariates per person in the cohorts. This is done using a sparse representation: covariates with a value of 0 are omitted to save space. The covariates object will have three columns: rowId, covariateId, and covariateValue. The rowId is usually equal to the person_id, unless specified otherwise in the rowIdField argument.

covariateRef An ffdf object describing the covariates that have been extracted.

metaData A list of objects with information on how the covariateData object was constructed.

---

getDbCovariateData *Get covariate information from the database*

Description

Uses one or several covariate builder functions to construct covariates.

Usage

```
getDbCovariateData(connectionDetails = NULL, connection = NULL,
  oracleTempSchema = NULL, cdmDatabaseSchema, cdmVersion = "5",
  cohortTable = "cohort", cohortDatabaseSchema = cdmDatabaseSchema,
  cohortTableIsTemp = FALSE, cohortId = -1, rowIdField = "subject_id",
  covariateSettings, aggregated = FALSE)
```

Arguments

connectionDetails	An R object of type connectionDetails created using the function createConnectionDetails in the DatabaseConnector package. Either the connection or connectionDetails argument should be specified.
connection	A connection to the server containing the schema as created using the connect function in the DatabaseConnector package. Either the connection or connectionDetails argument should be specified.
oracleTempSchema	A schema where temp tables can be created in Oracle.
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'.
cdmVersion	Define the OMOP CDM version used: currently supported is "5".
cohortTable	Name of the (temp) table holding the cohort for which we want to construct covariates
cohortDatabaseSchema	If the cohort table is not a temp table, specify the database schema where the cohort table can be found. On SQL Server, this should specifiy both the database and the schema, so for example 'cdm_instance.dbo'.

cohortTableIsTemp
: Is the cohort table a temp table?

cohortId
: For which cohort ID should covariates be constructed? If set to -1, covariates will be constructed for all cohorts in the specified cohort table.

rowIdField
: The name of the field in the cohort table that is to be used as the row_id field in the output table. This can be especially usefull if there is more than one period per person.

covariateSettings
: Either an object of type covariateSettings as created using one of the createCovariate functions, or a list of such objects.

aggregated
: Should aggregate statistics be computed instead of covariates per cohort entry?

Details

This function uses the data in the CDM to construct a large set of covariates for the provided cohort. The cohort is assumed to be in an existing table with these fields: 'subject_id', 'cohort_definition_id', 'cohort_start_date'. Optionally, an extra field can be added containing the unique identifier that will be used as rowID in the output.

Value

Returns an object of type covariateData, containing information on the covariates.

---

getDbDefaultCovariateData

Get default covariate information from the database

Description

Constructs a large default set of covariates for one or more cohorts using data in the CDM schema. Includes covariates for all drugs, drug classes, condition, condition classes, procedures, observations, etc.

Usage

```
getDbDefaultCovariateData(connection, oracleTempSchema = NULL,
  cdmDatabaseSchema, cohortTable = "#cohort_person", cohortId = -1,
  rowIdField = "subject_id", covariateSettings, aggregated = FALSE)
```

Arguments

connection
: A connection to the server containing the schema as created using the connect function in the DatabaseConnector package.

oracleTempSchema
: A schema where temp tables can be created in Oracle.

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'.

`cohortTable` Name of the table holding the cohort for which we want to construct covariates. If it is a temp table, the name should have a hash prefix, e.g. '#temp_table'. If it is a non-temp table, it should include the database schema, e.g. 'cdm_database.cohort'.

`cohortId` For which cohort ID should covariates be constructed? If set to -1, covariates will be constructed for all cohorts in the specified cohort table.

`rowIdField` The name of the field in the cohort temp table that is to be used as the row_id field in the output table. This can be especially usefull if there is more than one period per person.

`covariateSettings`
An object of type `defaultCovariateSettings` as created using the `createCovariateSettings` function.

`aggregated` Should aggregate statistics be computed instead of covariates per cohort entry?

Details

This function uses the data in the CDM to construct a large set of covariates for the provided cohort. The cohort is assumed to be in an existing temp table with these fields: 'subject_id', 'cohort_definition_id', 'cohort_start_date'. Optionally, an extra field can be added containing the unique identifier that will be used as rowID in the output. Typically, users don't call this function directly but rather use the `getDbCovariateData` function instead.

Value

Returns an object of type `covariateData`, containing information on the baseline covariates. Information about multiple outcomes can be captured at once for efficiency reasons. This object is a list with the following components:

covariates An ffdf object listing the baseline covariates per person in the cohorts. This is done using a sparse representation: covariates with a value of 0 are omitted to save space. The covariates object will have three columns: rowId, covariateId, and covariateValue. The rowId is usually equal to the person_id, unless specified otherwise in the rowIdField argument.

covariateRef An ffdf object describing the covariates that have been extracted.

metaData A list of objects with information on how the covariateData object was constructed.

`getDbHdpsCovariateData` *Get HDPS covariate information from the database*

Description

Constructs the set of covariates for one or more cohorts using data in the CDM schema. This implements the covariates typically used in the HDPS algorithm.

Usage

```
getDbHdpsCovariateData(connection, oracleTempSchema = NULL, cdmDatabaseSchema,
  cdmVersion = "4", cohortTable = "cohort_person", cohortId = -1,
  rowIdField = "subject_id", covariateSettings)
```

Arguments

connection	A connection to the server containing the schema as created using the connect function in the DatabaseConnector package.
oracleTempSchema	A schema where temp tables can be created in Oracle.
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'.
cohortTable	Name of the table holding the cohort for which we want to construct covariates. If it is a temp table, the name should have a hash prefix, e.g. '#temp_table'. If it is a non-temp table, it should include the database schema, e.g. 'cdm_database.cohort'.
cohortId	For which cohort ID should covariates be constructed? If set to -1, covariates will be constructed for all cohorts in the specified cohort table.
rowIdField	The name of the field in the cohort temp table that is to be used as the row_id field in the output table. This can be especially usefull if there is more than one period per person.
covariateSettings	An object of type covariateSettings as created using the createHdpsCovariateSettings function.
aggregated	Should aggregate statistics be computed instead of covariates per cohort entry?

Details

This function uses the data in the CDM to construct a large set of covariates for the provided cohort. The cohort is assumed to be in an existing temp table with these fields: 'subject_id', 'cohort_definition_id', 'cohort_start_date'. Optionally, an extra field can be added containing the unique identifier that will be used as rowID in the output. Typically, users don't call this function directly but rather use the getDbCovariateData function instead.

Value

Returns an object of type covariateData, containing information on the baseline covariates. Information about multiple outcomes can be captured at once for efficiency reasons. This object is a list with the following components:

covariates An ffdf object listing the baseline covariates per person in the cohorts. This is done using a sparse representation: covariates with a value of 0 are omitted to save space. The covariates object will have three columns: rowId, covariateId, and covariateValue. The rowId is usually equal to the person_id, unless specified otherwise in the rowIdField argument.

covariateRef An ffdf object describing the covariates that have been extracted.

metaData A list of objects with information on how the covariateData object was constructed.

getDefaultTable1Specifications *Get the default table 1 specifications*

Description

Loads the default specifications for a table 1, to be used with the `createTable1` function.

Usage

```
getDefaultTable1Specifications()
```

Value

A specifications objects.

loadCovariateData *Load the covariate data from a folder*

Description

`loadCovariateData` loads an object of type covariateData from a folder in the file system.

Usage

```
loadCovariateData(file, readOnly = FALSE)
```

Arguments

`file`	The name of the folder containing the data.
`readOnly`	If true, the data is opened read only.

Details

The data will be written to a set of files in the folder specified by the user.

Value

An object of class `covariateData`.

Examples

```
# todo
```

saveCovariateData *Save the covariate data to folder*

Description

saveCovariateData saves an object of type covariateData to folder.

Usage

```
saveCovariateData(covariateData, file)
```

Arguments

covariateData	An object of type covariateData as generated using getDbCovariateData.
file	The name of the folder where the data will be written. The folder should not yet exist.

Details

The data will be written to a set of files in the folder specified by the user.

Examples

```
# todo
```

tidyCovariateData *Tidy covariate data*

Description

Tidy covariate data

Usage

```
tidyCovariateData(covariateData, covariates, covariateRef, populationSize,
  minFraction = 0.001, normalize = TRUE, removeRedundancy = TRUE)
```

Arguments

covariateData	An object as generated using the getDbCovariateData function. If provided, the covariates, covariateRef, and populationSize arguments will be ignored.
covariates	An ffdf object with the covariate values in spare format. Will be ignored if covariateData is provided.
covariateRef	An ffdf object with the covariate definitions. Will be ignored if covariateData is provided. Only needed when removeRedundancy = TRUE.
populationSize	An integer specifying the total number of unique cohort entries (rowIds). Will be ignored if covariateData is provided. Only needed when removeRedundancy = TRUE.

`minFraction`	Minimum fraction of the population that should have a non-zero value for a covariate for that covariate to be kept. Set to 0 to don't filter on frequency.
`normalize`	Normalize the coviariates? (dividing by the max)
`removeRedundancy`	Should redundant covariates be removed?

Details

Normalize covariate values by dividing by the max and/or remove redundant covariates and/or remove infrequent covariates.

Index